# GENERAL POLICY FOR THE USE OF GENERATIVE AI IN ADMINISTRATIVE PROCESSES OF THE AEPD

## I. ANNEX IMPLEMENTATION OF THE AEPD'S GENERAL IAG POLICY

VERSION: 11 DECEMBER 2025

## A. USE CASES IN THE ADMINISTRATIVE SPHERE OF THE AEPD

The use cases presented do not exhaust the potential for the application of these technologies in the organisation, which may be progressively expanded and incorporated into this policy once identified.

The following table is intended as a guide and does not constitute an exhaustive or closed list of use cases, but rather an illustrative compilation of representative examples applicable within the framework of the policy. For each use case, the impact and risk are assessed qualitatively, evaluating the consequences of a use case failure and the level of control necessary to prevent it.

The table analyses the "Estimated institutional or functional impact" as an assessment of the relevance of the use case to the EDPS's activity and the institutional, operational or legal consequences for the EDPS or third parties if the use case fails or a threat materialises. It is rated as:

- Low, when it does not affect the achievement of the AEPD's objectives and there are no legal or significant consequences for the AEPD or third parties;
- Medium, when it could affect the quality or effectiveness of internal processes, but without legal or significant consequences for the AEPD or third parties;
- High, when it affects processes that may impact decisions or actions with legal, significant or reputational implications for the AEPD or third parties;

The column "Recommended system type" refers to one of these three types of systems:

| System types | Description |
|---|---|
| **External System** | Third-party IAG systems deployed on infrastructure outside the organisation's control, used as SaaS under their terms of use.<br>They are managed and maintained by external providers and are accessible via online platforms. In turn, they can be integrated into larger systems. They can be used to implement a phase in the procedure (services such as ChatGPT, Perplexity, MistralAI, Gemini, Claude, or others) or fully<br>integrated into the office environment, such as Microsoft 365 Copilot or Gemini Enterprise with Google Workspace. |
| **Internal system** | IAG systems developed by third parties and deployed on infrastructure controlled by the organisation.<br>The model is implemented on the organisation's own infrastructure or in a private cloud. Generally, but not limited to, using open source models such as ALIA, Llama, Qwen, Gemma, GPT-oss, Deepseek, Gemma, Phi, Kimi or Mistral. Also commercially licensed solutions that can be deployed entirely on the organisation's own infrastructure, including models (e.g. Mistral Enterprise On-premises).<br>premises). |
| **Ad-hoc system** | IAG systems developed internally or by third parties under custom specifications and deployed on infrastructure under the organisation's control and integrated with internal systems.<br>They offer the highest level of customisation and control. They include open-source models that undergo a fine-tuning process adapted to specific needs. |

The table also includes a column labelled "Specific observations/obligations". This column includes usage limitations or measures to be incorporated in each case.

| Use case | Description | Division/Sub-directorate Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Structuring, management or summarisation of general and open documents (public documents)** | Application of intelligent models to summarise, extract key information and transform public or non-confidential documents into structured formats that facilitate their analysis or reuse. Automated classification or analysis versions can be generated for study or dissemination purposes, ensuring anonymisation and the exclusion of personal data in all cases. exclusion of personal data. | All functional sub-directorates (varies by subject matter) | Low | External system | Editorial review No personal data or sensitive/confidential information. |
| **Translation of public documents** | Facilitate the translation of public documents into different languages. It allows documents written in foreign languages (e.g. resolutions from European authorities) to be translated into Spanish, or vice versa, adapting the style of the translation to the target audience. | All functional sub-directorates (varies by subject matter) | Low | External system | Editorial review No personal data or sensitive/confidential information. |
| **Generation of institutional communication content** | Comprehensive assistance in the production and review of institutional communication materials, campaigns and press releases, including the automated design of posters, brochures and corporate images<br>. | Press and Communications Office / General Secretariat | Under | External System | Editorial review No personal data or sensitive/confidential information. |

| Use case | Description | Division/Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Generation of diagrams, concept maps and infographics for communication** | Creation of diagrams, concept maps and explanatory infographics based on public information or open sources, to support the understanding and dissemination of content. | Press and Communications Office / General Secretariat | Under | External System | Editorial review No personal data or sensitive/confidential information. |
| **Generation of interactive multimedia and communication material and content** | Development of interactive multimedia content for education and public outreach, as well as automatic composition of music or ambient sound to accompany videos or public events<br>institutional or promotional events | Press and Communications Office / General Secretariat | Under | External System | Editorial review No personal data or sensitive/confidential information. |
| **Generation of audio or video files from open content and documentation** | Generation of synthetic voiceovers or videos, obtained exclusively from public or non-confidential documentation and without personal data, applicable to audio guides, telephone services or materials.<br>accessible. | Press and Communications Office / General Secretariat | Low | External System | Editorial review No personal data or sensitive/confidential information. |

| Use case | Description | Division/Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Development of informative or training content** | Support in the creation of internal or external awareness-raising and training materials on data protection, accessible and adapted to different profiles | Press and Communication Office / Technological Innovation Division / General Secretariat | Low | External System | Editorial review No personal data or sensitive/confidential information. |
| | Generation of content and news on the institutional website and social media. | | Low-Medium | External System | Editorial review No personal data or sensitive/confidential information. |
| **Identification of trends or patterns from open sources open sources** | Identification of trends or patterns through automated analysis of open information to support decision-making. | All functional sub-directorates (varies by subject matter) | Low | External system | Editorial review No personal data or sensitive/confidential information. |

| Use case | Description | Division/Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Preparation of graphs, tables and reports based on publicly available or non-confidential information** | Automatic preparation of financial or budgetary graphs, tables and reports based on publicly available or non-confidential data, ensuring the exclusion of personal or sensitive information. | All functional sub-directorates (varies by subject matter) | Low | External system | Editorial review No personal data or sensitive/confidential information. |
| **Assistance in the general approach and preparation of arguments, legal and technical studies** | Use of generative AI systems for the analysis and comparison of legal, doctrinal or technical sources of a public or non-confidential nature, with the aim of developing argument frameworks, comparative studies, technical notes or interpretative hypotheses. These tools support the preparation of legal strategies or approaches without replacing professional assessment or implying automated decision-making.<br>They can also be used for exploratory and systematic analysis of legal or open sources, comparison of regulatory or doctrinal frameworks, and preparation of general studies that do not involve confidential information or personal data<br>. | All functional sub-directorates (varies by subject matter) | Under | External System | Editorial review No personal data or sensitive/confidential information. |

| Use case | Description | Division/Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Assistance in the initial drafting of general and isolated elements whose ideas can then be incorporated into resolutions, reports or technical guidelines** | Use of generative AI systems for the analysis and comparison of public or non-confidential legal, doctrinal or technical sources, in order to develop argumentative frameworks, comparative studies, technical notes or interpretative hypotheses. These tools support the preparation of legal strategies or approaches without replacing professional assessment or implying automated decision-making.<br>They can also be used for the exploratory and systematic analysis of legal or open sources, the comparison of regulatory or doctrinal frameworks, and the preparation of general studies that do not involve confidential information or personal data<br>. | All functional sub-directorates (varies by subject matter) | Low | External system | Editorial review No personal data or sensitive/confidential information. |
| **Assistance with system development and configuration tasks** | AI has become a key tool for technical tasks such as scripting, generating SQL queries, building regular expressions, and resolving questions about specific software.<br>It facilitates much of the work, allowing users to focus<br>on review and fine-tuning rather than | General Secretariat | Low-Medium | External/Internal System | Editorial review No personal data, sensitive/confidential information, or information about corporate system architectures and configurations. |

| Use case | Description | Division / Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| | Start from scratch. This is especially useful in environments where multiple technologies and advanced system configurations are managed. | | | | |
| **Internal regulatory or doctrinal assistant** | Development of internal virtual assistants that facilitate quick consultation of legislation, resolutions, legal reports, the Agency's interpretative criteria, etc., improving access to organisational knowledge. They can be used to support inspections, the various stages of processing applications for the approval of codes of conduct and applications for international transfer agreements, or the preparation of anonymised summaries to support the dissemination of content, among other things. | All functional sub-directorates (varies by subject matter) | Low (if only public documents are used) -Medium | External system (if only public documents are used) Internal system / Ad-hoc system | Editorial review.<br>No personal data or sensitive/confidential information.<br>Must not generate binding content.<br>Periodic updating of document corpus. |
| | | Deputy Directorate-General for Promotion and Authorisations | Medium | Internal system / Ad-hoc system | Editorial review and human control if decisions are involved.<br>Must not generate binding content.<br>Periodic updating of document corpus. |
| | | Press and communications | Media | Internal system / Ad-hoc system | Editorial review.<br>No personal data or sensitive/confidential information.<br>It must not generate binding content binding content. Regular updating of the document corpus. |

| Use case | Description | Division/Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
|  |  | General Sub-Directorate of Inspection | Medium-high if it uses personal data or sensitive/confidential information, or involves decisions that may impact third parties. | Internal system / Ad-hoc System | Editorial review and human control if it involves decisions.<br>It should not generate binding content. Periodic updating of the document corpus. |
| **Transcription of audio and video from open sources without confidential information confidential** | Transcription of multimedia content from open or non-confidential sources for incorporation into different workflows or studies. May contain personal data. | All functional sub-directorates (varies by subject matter) | Medium | Internal system | Editorial review.<br>No sensitive/confidential information.<br>Must not generate binding content. |
| **Transcription of audio and video containing internal, private or confidential information, where applicable, from meetings and support for the generation of minutes** | Transcription of internal meetings, where applicable, to support the generation of minutes. As well as specific audio recordings of interviews granted or speeches made by Agency staff at events and the preparation of summaries of the resulting text. | All functional sub-directorates (varies by subject) Press and communication | Media | Internal system | Editorial review and human control if decisions are involved. Must not generate binding content. Periodic updating of the document corpus. |

| Use case | Description | Division/Sub-directorate<br>Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Structuring, management or summaries of internal administrative documents or documents containing personal data** | Application of intelligent models to summarise, extract key information and convert internal administrative documents into standardised structures that improve their automated processing or analysis. In these cases, complete or partial anonymisation of personal data must be applied and processing must be limited to secure and authorised environments, for example, to produce anonymised versions for transparency queries<br>transparency or internal reports. | All functional sub-directorates (varies by subject matter) | Medium<br>Personal or confidential/sensitive data. | Internal system / Ad-hoc system | Editorial review and human control if decisions are involved. Must not generate binding content. Periodic updating of document corpus. |
| **Drafting of letters, emails, or internal memos with an institutional tone institutional tone** | Use of intelligent systems to draft initial responses to citizens or organisations, or written communications, speeding up processing times and ensuring consistency in<br>institutional messages. | General Secretariat | Medium | Internal System / Ad-hoc System | Editorial review and human control if decisions are involved. It should not generate binding content. Periodic updating of the document corpus. |
| **Classification and summary of complaints, claims, queries and other entries.** | The use of RAGs and well-tuned systems will allow access to real-time updated information on all material produced by the AEPD to produce draft responses to specific queries from the citizen service channels, the DPO channel and the youth channel.<br>youth channel. | Deputy Directorate-General for Promotion and Authorisations | Medium-high | Internal system/Ad-hoc system | Editorial review and human control if decisions are involved.<br>Must not generate binding content. Periodic updating of document corpus. |

| **Use case** | **Description** | **Division/Sub-directorate Main** | **Estimated institutional or functional impact** | **Recommended system type** | **Observations / Specific obligations** |
|---|---|---|---|---|---|
| **Classification and summary of complaints, claims, enquiries and other entries.** | Application of natural language processing (NLP) models to facilitate the initial processing of information received by the Agency, enabling its categorisation, indexing and preliminary analysis.<br>. | Deputy Directorate-General for Promotion and Authorisations | Environment | Internal System/Ad-Hoc System | Editorial review and human control if decisions are involved. It should not generate binding content. Periodic updating of the document corpus. |
| **Support in the analysis of the EIPD (Data Protection Impact Assessment)** | Use of expert models to perform an initial triage or provide documentary support in impact assessment reports, providing templates, common criteria or analysis guidelines. | Deputy Directorate-General for Data Inspection/Technological Innovation Division/DPD | Medium-high | Internal system / Ad-hoc System | Editorial review and human control if decisions are involved. Should not generate binding content. Periodic updating of the document corpus. |
| **Intelligent management of the strategic plan** | Automating strategic plan management facilitates real-time monitoring of indicators related to the fulfilment of objectives and results. It allows for the automation of the chosen methodology, optimises strategy execution, and enables data-driven decision-making in real time. Data analysis and visualisation of results for monitoring plans and indicators. | Presidency/Deputy | Senior | Internal System / Ad-hoc System | Editorial review and human oversight if decisions are involved. |

| Use case | Description | Division/Sub-department Main | Estimated institutional or functional impact | Recommended system type | Observations / Specific obligations |
|---|---|---|---|---|---|
| **Assistance in drafting resolutions, reports or technical guidelines.** | Use of generative AI tools to support the initial drafting of regulatory and doctrinal documents, facilitating content structuring and increasing efficiency in document production. | Deputy Directorate-General for Data Inspection/Technological Innovation Division/DPD | Medium-high | Internal System / Ad-hoc System | Editorial review and human control if decisions are involved.<br>Must not generate binding content. Periodic updating of document corpus. |
| **Intelligent alert systems (prioritisation of reports, complaints, notifications and other entries)** | Development of AI or RPA systems that enable the detection and prioritisation of reports, data breaches or sensitive communications, including automatic alerts for high-impact cases, vulnerable groups or situations that require priority attention at the Agency's discretion, thus enabling agile and focused management. | Deputy Directorate-General for Data Inspection/Technological Innovation Division/Press and Communications Office | High | Internal System / Ad-Hoc System | Editorial review and human control if decisions are involved. Must not generate binding content. Periodic updating of document corpus. |
| **Support for the processing of files and notifications of data breaches** | The processing of files and/or notifications of personal data breaches are manual processes assisted by various corporate tools. IAG speeds up tasks, from the initial classification of entries by record, their categorisation and extraction of structured information, to the preparation of summaries and drafts. | Deputy Directorate-General for Data Inspection/Technological Innovation Division | High | Internal System/Ad-Hoc System | Editorial review and human control if decisions are involved. It should not generate binding content. Periodic updating of the document corpus. |

## B. DETAILED ANALYSIS OF THREATS IN RELATION TO THE IAG SYSTEM IMPLEMENTED IN THE PROCESSES.

Below are representative examples of threats identified in the application of generative artificial intelligence (GAI) systems to administrative processes, based on the use cases analysed by the AEPD. These threats do not constitute an exhaustive analysis, but rather an initial reference that will be adapted or expanded as necessary. These threats are graded according to the type of GAI system implemented in the processes:

- External System (third-party GAI systems deployed on infrastructure outside the organisation's control, used as SaaS under their terms of use)
- Internal System (IAG systems developed by third parties and deployed on infrastructure under the organisation's control)
- Ad-hoc system (IAG systems developed internally or by third parties under custom specifications and deployed in infrastructure under the organisation's control and integrated with internal systems

The purpose of the general policy proposed by the AEPD is to adequately manage these threats in order to either reduce their impact or reduce the likelihood of their materialisation through the appropriate selection of IAG systems for specific use cases, the application of organisational and technical measures, and the implementation of appropriate governance oversight mechanisms and effective policy management.

The following tables describe examples of technical and operational threats that may affect generative artificial intelligence (GAI) systems, classified according to the type of system (external, internal or ad hoc). The High/Medium/Low ratings do not represent a probability or quantitative impact, but rather a qualitative estimate of the system's level of exposure or vulnerability to each threat, based on the degree of control that the organisation can exercise over the model, data and operating environment.

### 1. Threats to the effectiveness of GAI systems

The effectiveness of AI systems affects, among other things, the protection of fundamental rights, insofar as the competence of the AEPD is precisely the protection of those rights in relation to the processing of personal data.

| Threat | External System | Internal System | Ad-hoc System |
|---|---|---|---|
| Hallucinations: Plausible but false or invented responses | **High**: no control over the model, difficult to detect systematic errors. | **Medium**: possibility of incorporating review mechanisms. | **Low**: adapted and/or trained with controlled corpus and designed for the domain. |
| Lack of specific knowledge: Inability of the model to respond accurately in specialised areas. | **High**: generalist models, trained for broad contexts. | **Medium**: can be connected to internal sources. | **Low**: adapted and/or trained for use with data from the AEPD environment. |
| Irrelevant correlations | **Medium**: opaque model, with no possibility of adjusting behaviour. | **Medium**: partial control, input/output parameters can be adjusted. | **Low**: more rigorous adaptation and/or training design, lower risk of logical errors. |
| Non-repeatable outputs: Variability in responses to identical inputs, hindering consistency, due to the lack of control over the context accessed by the model | **High**: unpredictable behaviour in each API call. | **Low**: model configuration/parameterisation under control. | **Low**: model configuration/parameterisation under control, possibility of standardising outputs. |
| Accelerated technological obsolescence, especially if immature solutions or those with a strong dependence on the provider are adopted. on the supplier | **High**: market solutions may become outdated or abandoned. | **Medium**: depends on the model's life cycle and internal technical capacity. | **Low**: design adapted to the institutional context with the possibility of planned evolution . |

## 2. Threats of bias and discrimination

| Threat | External System | Internal System | Ad-hoc System |
|---|---|---|---|
| Biases in training data: Reduction in critical thinking and human validation. | **High**: the origin is unknown and cannot be audited. The model may carry cultural or demographic biases. | **Medium**: possibility of reviewing part of the model or applying filters, although the original dataset is not always accessible. | **Low**: the dataset can be selected, curated and audited to ensure diversity, balance and representativeness. |
| Algorithmic biases | **High**: it is not possible to modify the architecture or know in detail the algorithmic biases incorporated. | **Medium**: minor adjustments (retraining, fine-tuning) can be applied, but the base model design is not accessible. | **Low**: the model design, hyperparameter selection and training are in the hands of the organisation. The introduction of bias can be reduced from the outset. |

## 3. Impacts on rights and freedoms in relation to data protection

This set of threats applies primarily when the IAG system is used in AEPD processes involving personal data, such as those derived from use cases in which texts containing information about individuals are processed.

There are many use cases in which the use cases would not explicitly involve the processing of personal data, such as in the case of analysis or summaries of regulations. In these cases, most of which have a lower impact, it is necessary to take into account the threats posed by the collection of metadata and profiling of the users of IAG systems themselves, or the manipulation and theft of information from the users themselves through indirect injection techniques (prompts).

| Incident or risk scenario | Threats according to LIINE4DU methodology[1] | External System | Internal System | Ad-hoc System |
|---|---|---|---|---|
| Leaks or misuse of personal data**:** from other members of the organisation or third parties. | Link Identification<br>Data breach<br>Disclosure | **Very high**: risk of reuse for training, lack of contractual guarantees, records on supplier servers provider's servers. | **Low**: the environment is controlled by the AEPD; ENS measures and internal policies can be applied. | **Very low**: the entire processing cycle is under institutional supervision, with no external transfer. |
| Exposure of personal data during training or use | Linking<br>Identification Data breach Disclosure | **High**: data sent via prompts may be recorded or reused by the provider. | **Low**: if the data is used in a closed environment without an external connection. | **Very low**: data is selected and managed securely; possibility of prior anonymisation. |
| Profiling of the user (natural person) of the IAG through the collection of prompts, metadata or access to personal information stored in the system | Linking<br>Identification<br>Detection<br>Deception | **Very high**: external providers' terms of use may include retraining or input analysis. | **Low**: in a local environment, processing is limited by the technical policies themselves. | **Non-existent**: data does not leave the institutional environment, nor is it shared or reused externally. |
| Difficulty in deleting data relating to individuals processed in interaction with the service. | Non-repudiation Lack of knowledge and inability to intervene | **Very high**: there are no guarantees of effective deletion or control over backups or retrained models. | **Medium**: depends on the storage and configuration of the local model. | **Low**: possibility of using reversible techniques (LoRA, Adapters) and applying controlled deletion. |
| Unauthorised access to internal documents or sources containing personal data | Data breach<br>Disclosure | **High**: greater attack surface attack surface and less control over access and storage. | **Medium**: mitigated risk through technical measures (ENS, roles, encryption). | **Low**: complete control over the environment and access policies. |
| Lack of traceability and auditing of data used | Inaccuracy Lack of knowledge and inability to intervene | **Very high**: total opacity of the model and the provider's internal processes. | **Medium**: partial traceability possible if local processes are documented. | **Low**: traceability guaranteed if defined from the design. |

[1] Introduction to liine4du 1.0: a new methodology for modelling threats to privacy and data protection

| | | | | |
|---|---|---|---|---|
| Persistence of personal data in logs held by the service provider or third party | Linking<br>Identification Data breach<br>Disclosure | **High**: logs may be kept outside the control of the AEPD. | **Medium**: requires internal control of records and periodic disposal. | **Low**: can be anonymised or purged automatically. |
| Loss of control over pre-trained models | Inaccuracy Exclusion | **Very high**: impossible to know the data used or modify the model. | **Medium**: lower risk, although pre-training still cannot be audited. | **Low**: model fully traceable and trained with known data. |
| Vulnerabilities in interfaces, APIs, or adversarial attacks or backdoors | Linking<br>Identification Data breach<br>Disclosure | **High**: high exposure in online environments, open or unaudited APIs. | **Medium**: mitigable with good cybersecurity practices. | **Low**: smaller attack surface and closed design according to ENS. |
| Physical security of protected persons if information about schedules, movements or sensitive decisions is exposed | Link Identification Data breach<br>Disclosure | **Low**: there would be a risk if sensitive operational data or agendas were used in accessible systems, which is not the case with the AEPD. | **Low**: this could be mitigated by segmenting access and shielding data, but this is not the case with the AEPD. | **Low**: complete control of the environment and exclusion of sensitive information. |

## 4. Threats to infrastructure security and business continuity

| Incident or risk scenario | Threats according to STRIDE[2] | External system | Internal System | Ad-hoc System |
|---|---|---|---|---|
| Technical vulnerabilities in systems hosting IAG systems | Identity theft<br>*Tampering*<br>Denial of service | **High**: exposed to external vectors, without full control over patches or security layers security layers. | **Medium**: mitigable with ENS measures, segmentation and hardening. | **Low**: closed environment, designed and secured according to internal standards. |
| Uncontrolled internet access, with risk of external threats. | *Tampering*<br>Information disclosure | **High**: risk of malware intrusion, connection to sources | **Medium**: if access is restricted access and properly monitored. | **Low**: can be completely isolated if necessary. |

[2] OWAST Threat Modelling Process

| | | | | |
|---|---|---|---|---|
| | Privilege escalation | insecure, downloading malicious content. | | |
| Prompt injection or adversarial attacks | *Tam*pering<br>Information disclosure<br>Privilege escalation | **High**: external providers are more susceptible to sophisticated attacks without direct supervision. | **Medium**: mitigable with input validation and interface control. | **Low**: allows for protection by design and total input control. |
| Human errors in technical management or environment configuration. | Identity theft<br>*Tampering*<br>Repudiation Disclosure of information<br>Denial of service<br>Privilege escalation | **High**: dependence on third parties, lack of knowledge of the backend. | **Medium**: manageable with training and internal procedures. | **Low**: complete control by specialised personnel and specific training. |
| Resilience and availability: Risk of service interruption due to lack of support, licences or discontinuity of the product | Denial of service | **High**: risk of unilateral changes to conditions, cancellation of services, lack Support or licences. | **Medium**: will depend on technical maintenance and internal support. | **Low**: continuity depends on the internal evolution and scalability plan. scalability plan. |
| System portability: Difficulty in migrating or replacing the system due to lack of interoperability or technological dependency | Denial of service | **High**: limited control over format, code and model. Difficulty in migrating without loss. | **Medium**: greater flexibility if the entire technical cycle is managed. | **Low**: architecture designed to be reusable or migrable. |
| Possible exposure of sensitive information linked to individuals or senior institutional officials | Identity theft<br>Disclosure of information | **High**: serious risk if integrated into open environments or uncontrolled systems. | **Medium**: can be protected with encryption, roles and segmentation. | **Low**: only accessible from secure channels defined by the AEPD. |

## 5. Disclosure of non-personal information (sensitive/confidential information on the AEPD's actions)

| Threat | External system | Internal System | Ad-hoc System |
|---|---|---|---|
| Institutional disclosure when exposing internal criteria, strategies or approaches that may affect the image or authority of the AEPD | **Very high**: risk of memorising prompts or responses containing critical institutional information without guarantees of deletion or confidentiality. | **Medium**: limited control if closed models are reused. | **Low**: fully supervised content and use; no risk of exposure outside the environment. |
| Disclosure of strategic interests by revealing priority areas of action that could be used by third parties to anticipate. | **High**: the content generated may indirectly expose priority lines or sensitive issues sensitive topics. | **Medium**: risk if inputs/outputs and indexed corpus are not controlled. | **Low**: possibility of defining rules for excluding sensitive topics. |
| Revelation of actions by third parties subject to inspection through AI-generated content that allows for the deduction or identification of procedures that are not yet public | **Very high**: the model may return inferred information that indirectly links to actual past or ongoing actions progress. | **Medium**: mitigable with source and prompt control. | Low: supervision of content and generated outputs. |
| Disclosure of internal working documents such as drafts, minutes or exchanges that compromise organisational confidentiality | **High**: high risk if drafts, minutes, or confidential texts are used in prompts or as Sources in external RAG systems. | **Medium**: can be prevented with access segmentation and document control. | **Low**: documents are kept in internal repositories with controlled access management. |
| Poisoning of data sources to manipulate responses, actions, and theft of corporate information (prompting attacks indirect attacks) | **High**: if the RAG system accesses online sources, it could incorporate manipulated or false data. | **Medium**: may occur if the internal corpus is not properly filtered. | **Low**: corpus curated, validated and managed by the organisation. Control over the ingestion cycle. |
| Filtering of relationships with other authorities or public/private entities, affecting institutional cooperation or perceived neutrality perceived | **High**: possibility of revealing collaborations, exchanges or joint actions that are not published. | **Medium**: depends on the corpus used and the logs generated. | **Low**: adaptation and/or training and use are under strict functional and legal supervision. |

## 6. Disclosure of non-personal (sensitive/confidential third-party) information

| Threat | External system | Internal system | Ad-hoc System |
|---|---|---|---|
| Disclosure of trade secrets provided in proceedings or consultations, when used as examples or prompts without protection | **Very high**: possibility of retention and uncontrolled use of sensitive examples or content in the prompts. | **Medium**: risk if used without labelling or isolation in the corpus. | **Low**: controlled adaptation and/or training with explicit exclusion of classified information. |
| Exposure of confidential technical or legal information included in documentation submitted by companies, organisations or advisors | **High**: external models could record or infer patterns of sensitive content. | **Medium**: mitigable if filters and document restrictions are applied. | **Low**: possibility of excluding categories or flagging confidential documents in the system design. |
| Filtering of elements protected by intellectual property such as software, algorithms or models, if reused as part of training | **Very high**: legal risks if software, business logic or technical documentation is incorporated into models without control. | **Medium**: requires strict control of data sources and licences. | **Low**: adaptation and/or training with authorised materials or materials generated by the AEPD. |
| Loss of institutional trust by third parties when they perceive a risk of reuse or exposure of the information provided. | **High**: the use of commercial platforms may generate mistrust regarding the protection of the information provided. | **Medium**: if traceability and document protection are not guaranteed. | **Low**: control of the environment and documentation of guarantees offered to third parties. |
| Unauthorised disclosure of content marked as confidential by other organisations, companies or entities in the public sector sector | **High**: possible reuse or inference of content marked as confidential by other organisations. | **Medium**: mitigable with internal classification and segmentation policies. | **Low**: strict compliance with clauses and control of the data cycle. |

## 7. Incorrect, irresponsible or harmful human interaction with the IAG

| Threat | External System | Internal System | Ad-hoc System |
|---|---|---|---|
| Failure to comply with AI system usage policies regarding the use of personal, confidential or sensitive information sensitive information. | **High:** No measures by design that prevent | **Low:** impossible due to measures by design | **Low:** impossible due to measures from the design stage |
| Negative impact on staff working conditions working conditions for staff, in the event that the | **Medium:** not dependent on the | | |

| | |
|---|---|
| implementation of IAG in processes involves the objectification of employees or pressure for increased productivity beyond<br>beyond the rational use of IAG | |
| Resistance to change | **Medium:** not dependent on the system |
| Perception of IAG as a threat:<br>as excessive or undue monitoring of work performance through IAG systems. | **Medium:** not dependent on the system |
| Incorrect use of tools with<br>loss of effectiveness | **Means:** not dependent on the system |
| Uncritical interaction with intelligent systems can be a source of automated decisions, errors, biases, or inappropriate decisions.<br>inappropriate decisions | **High:** not dependent on the system |

## 8. Lack of transparency and explainability of actions based on AI or lack of consistency in similar situations or deviations in the application of current criteria

| Threat | External System | Internal system | Ad-hoc system |
|---|---|---|---|
| Perception of objectification and dehumanisation in processes<br>citizen-oriented processes | **Medium:** Does not depend on the system | | |
| Reproduction or amplification of social inequalities if the system indirectly discriminates against vulnerable groups | **High**: external training may contain uncorrected biases. | **Medium**: partial control if the dataset is reviewed or fine-tuning is applied. | **Medium**: possible to eliminate biases from the adaptation and/or selection of the corpus and the training criteria. |
| Loss of public confidence in the AEPD if decisions or actions are perceived to be influenced by opaque automated systems | **Very high**: the use of opaque commercial platforms can erode the perception of<br>impartiality. | **Medium**: if the scope of AI use is not adequately communicated. | **Low**: high traceability, transparency and visible governance. |

| | | | |
|---|---|---|---|
| Social impact of a breach | **High:** due to data communication and dependence on external systems. | **Low:** the external process is minimised. | **Low:** the external process is minimised. |
| Lack of explainability | **High:** generally, no explainability information is provided and the tests performed by the user do not control all parameters. | **Medium:** generally, no explainability information is provided, but there is greater test control. | **Low:** It can be controlled and explainability information can be obtained. |
| Excessive user confidence | **High**: The opacity of the external system and its appearance of authority increase the risk of uncritical acceptance. | **Medium**: greater control over the environment may encourage more conscious use, but there may still be excessive confidence. | **Low**: as part of a controlled institutional process, it is easier to implement training, validation and systematic review. |
| Failures in the operation of the AI system | **High:** there is less possibility of controlling the effectiveness of systems | **Medium:** there is a greater possibility of controlling the effectiveness of the systems | **Medium:** there is a greater possibility of controlling the effectiveness of the systems |

## 9. Misgovernment and loss of institutional integrity

| Threats | External System | Internal System | Ad hoc system |
|---|---|---|---|
| Fraud or manipulation of procedures | **High**: greater exposure if internal logic is unknown and system behaviour cannot be audited of the system. | **Medium**: mitigable with cross-validations and local logs. | **Low**: design based on internal control and traceability principles. |
| Internal financial impact due to maintenance, licences, scalability or unexpected cost overruns. | **High**: licences, variable prices, technological lock-in and low long-term predictability. | **Medium**: controllable costs, may be dependent on updates or external support. . | **Medium**: higher initial investment, but predictable economic sustainability. |
| Loss of institutional control over critical functions if excessively delegated to | **Very high**: if relevant decisions are delegated to opaque or third-party models. | **Medium**: mitigable with validation and traceability. | **Low**: decisions remain within the institutional sphere. |

| | | | |
|---|---|---|---|
| technologies without reversibility or auditability | | | |
| Dysfunctions in inter-administrative coordination if common standards are not followed or systems without interoperability | **High**: risk if solutions are used that are not interoperable or not aligned with AGE standards. | **Medium**: can be aligned with common technology policies. | **Low**: design compliant with interoperability criteria and public standards. |

## C. IMPROVEMENT OF RESULTS AND ADAPTATION TO THE DOMAIN OF USE

To improve the performance and adaptation of models to specific needs, two techniques are used, which can be complemented with reinforced learning techniques, such as:

- Fine-tuning or specific adjustment: Starting from a pre-trained model, its parameters (some or all) are readjusted through retraining on a specific data set. This process allows an LLM to be adapted to specific terminology, style and requirements.
- Retrieval-Augmented Generation (RAG): The model is not modified, but there is a preliminary phase of real-time information *retrieval* from databases, documents, the internet, or other sources. This retrieved information is added to the LLM input *prompt*. This allows access to updated information without the need to retrain the model.

## D. DEPLOYMENT PLAN

A structured plan to achieve the objectives of the governance model will need to include the following milestones (which may overlap):

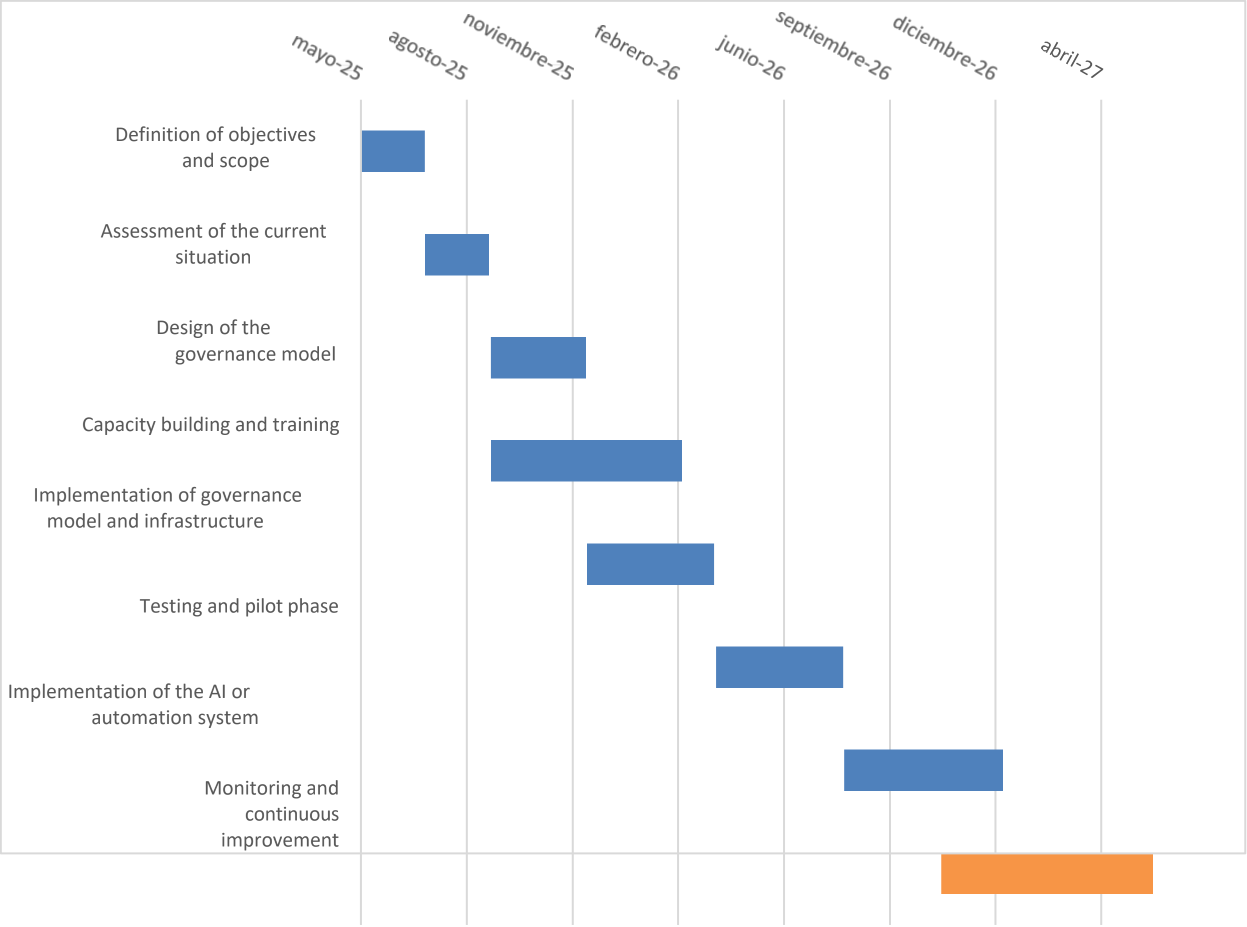
- Definition of objectives and scope: Establish the specific objectives of the governance model and determine its scope, identifying the areas and processes where IAG can generate the most value and which will be covered. Involve staff and users in identifying needs (1 to 2 months).
- Assessment of the current situation: Identify and catalogue available data, existing procedures and processes, and technological infrastructure, detecting potential problems. Assess the feasibility and impact of the IAG solution (2 months).
- Design of the governance model: Define roles and responsibilities, establish the corresponding policies and procedures (3 months).
- Capacity building and training: Train the staff involved and raise awareness among the rest of the staff. Organise training courses and workshops, support materials and practical guides (4 to 6 months).
- Implementation of the governance model and infrastructure: Implement the defined policies and procedures. Implement the technical infrastructure. Formalise contracts with suppliers and managers. Integrate monitoring and audit measures (4 to 6 months).
- Testing and pilot phase: implement prototypes of the selected IAG solution, evaluating and refining its performance first in a

[3] See, for example, the guide published by the CCN for creating a chatbot that uses third-party models locally with the option of also including RAG (https://www.ccn-cert.cni.es/es/seguridad-al-dia/novedades-ccn-cert/13063-como-crear-un-chatbot-con-llm-de-forma-local.html).

tests and then a pilot in a real environment, in a selection of possible use cases. Adjust and optimise the system (3 to 6 months, longer if an ad-hoc model is used).

- Implement the IAG system in all use cases (3 to 6 months or more, depending on the number of use cases and their size).
- Monitoring and continuous improvement: Continuous evaluation of the governance model. Updating policies and procedures as necessary. Keeping detailed records of activities. Expanding use cases and extrapolating to other processes. Conducting audits (every 3 or 4 months, at least annually).

Below is an example of provisional planning:


mayo-25
agosto-25
noviembre-25
febrero-26
junio-26
septiembre-26
diciembre-26
abril-27
Definition of objectives and scope
Assessment of the current situation
Design of the governance model
Capacity building and training
Implementation of governance model and infrastructure
Testing and pilot phase
Implementation of the AI or automation system
Monitoring and continuous improvement