POLÍTICA GENERAL PARA EL USO DE IA GENERATIVA EN PROCESOS ADMINISTRATIVOS DE LA AEPD

I. ANEXO IMPLEMENTACIÓN DE LA POLÍTICA GENERAL IAG DE LA AEPD

VERSIÓN: 11 DE DICIEMBRE DE 2025

A. CASOS DE USO EN EL ÁMBITO ADMINISTRATIVO DE LA AEPD

Los casos de uso planteados no agotan el potencial de la aplicación de estas tecnologías en la organización, que podrá ampliarse progresivamente y que se incorporarán a la presente política una vez se identifiquen.

La siguiente tabla tiene carácter orientativo y no constituye una lista exhaustiva ni cerrada de casos de uso, sino una recopilación ilustrativa de ejemplos representativos aplicables en el marco de la política. Para cada caso de uso se evalúan el impacto y el riesgo de manera cualitativa, valorando las consecuencias que tendría un caso de uso que falla y el nivel de control necesario para evitarlo.

En la tabla se analiza el "Impacto institucional o funcional estimado" como la evaluación de la relevancia del caso de uso para la actividad de la AEPD y las consecuencias institucionales, operativas o jurídicas, para la AEPD o para terceros, si el caso de uso falla o se materializa una amenaza. Se califica como:

- Bajo, cuando no afecta a la consecución de los objetivos de la AEPD y no existen consecuencias jurídicas o significativas para la AEPD o terceros;
- Medio, cuando podría afectar a la calidad o eficacia de procesos internos, pero sin consecuencias jurídicas o significativas para la AEPD o terceros;
- Alto, cuando incide en procesos que pueden afectar a decisiones o actuaciones con implicaciones jurídicas, significativas o reputacionales para la AEPD o terceros;

La columna "Tipo de sistema recomendando" hace referencia a uno de estos tres tipos de sistemas:

Tipos de sistemas	Descripción
Sistema Externo	Sistemas IAG de terceros desplegados en infraestructura fuera del control de la organización, utilizados como SaaS bajo sus términos de uso. Están gestionados y mantenidos por proveedores externos y están accesibles a través de plataformas en línea. A su vez, pueden estar integrados en sistemas más amplios. Pueden ser utilizados para implementar una fase en el procedimiento (servicios como ChatGPT, Perplexity, MistralAI, Gemini, Claude u otros) o plenamente integrados en el entorno ofimático como Microsoft 365 Copilot o Gemini Enterprise con Google Workspace.
Sistema interno	Sistemas IAG desarrollados por terceros y desplegados en infraestructura bajo control de la organización. El modelo se implementa en la infraestructura propia o en una nube privada. Generalmente, aunque no limitado, utilizando modelos de pesos abiertos como ALIA, Llama, Qwen, Gemma, GPT-oss, Deepseek, Gemma, Phi, Kimi o Mistral. También soluciones bajo licencia comercial que se puedan desplegar íntegramente en infraestructura propia de la organización, incluyendo los modelos (Por ejemplo, Mistral Enterprise On-premises).
Sistema Ad-hoc	Sistemas IAG desarrollados internamente o por terceros bajo especificaciones a medida y desplegados en infraestructura bajo control de la organización e integrado con sistemas internos. Ofrecen el máximo nivel de personalización y control. Incluyen modelos de código abierto sobre los que se realiza un proceso de fine-tuning adaptado a necesidades específicas.

En la tabla también aparece una columna etiquetada como “Observaciones/Obligaciones específicas”. En dicha columna se incluyen limitaciones de uso o medidas a incorporar para cada caso.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Estructuración, gestión o resumen de documentos generales y abiertos (documentos públicos)	Aplicación de modelos inteligentes para resumir, extraer información clave y transformar documentos de carácter público o no confidencial en formatos estructurados que faciliten su análisis o reutilización. Pueden generarse versiones de clasificación o análisis automatizado con fines de estudio o divulgación, garantizando en todo caso la anonimización y la exclusión de datos personales.	Todas las Subdirecciones funcionales (varía por materia)	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Traducción de documentos públicos	Facilitar la traducción de documentos públicos en diferentes idiomas. Permite traducir documentos redactados en lenguas extranjeras (por ejemplo, resoluciones de autoridades europeas) al español, o viceversa, adaptando el estilo de la traducción según el público objetivo.	Todas las Subdirecciones funcionales (varía por materia)	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Generación de contenidos de comunicación institucional	Asistencia integral a la producción y revisión de materiales de comunicación institucional, campañas y notas de prensa, incluyendo el diseño automatizado de carteles, folletos e imágenes corporativas.	Gabinete de Prensa y Comunicación / Secretaría General	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Generación de esquemas, mapas conceptuales e infografías para la comunicación	Creación de esquemas, mapas conceptuales e infografías explicativas elaboradas a partir de información pública o de fuentes abiertas, para apoyar la comprensión y difusión de contenidos.	Gabinete de Prensa y Comunicación / Secretaría General	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Generación material y contenidos multimedia interactivo y de comunicación	Desarrollo de contenido multimedia interactivo destinado a la educación y divulgación ciudadana, así como composición automática de música o sonido ambiental para acompañar vídeos o actos públicos institucionales o promocionales	Gabinete de Prensa y Comunicación / Secretaría General	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Generación de audios o vídeos a partir de contenidos y documentación abierta	Generación de locuciones o vídeos sintéticos, obtenidas exclusivamente a partir de documentación pública o no confidencial y sin datos personales, aplicables a audioguías, servicios telefónicos o materiales accesibles.	Gabinete de Prensa y Comunicación / Secretaría General	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Elaboración de contenidos divulgativos o formativos	Apoyo en la creación de materiales de sensibilización y formación interna o externa sobre protección de datos, accesibles y adaptados a distintos perfiles	Gabinete de Prensa y Comunicación / División de Innovación Tecnológica / Secretaría General	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
	Generación de contenidos y noticias en la web institucional y redes sociales.		Bajo-Medio	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Identificación de tendencias o patrones a partir de fuentes abiertas	Identificación de tendencias o patrones mediante el análisis automatizado de información abierta para el apoyo a la toma de decisiones.	Todas las Subdirecciones funcionales (varía por materia)	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Elaboración de gráficas, tablas e informes a partir de información de acceso público o no confidencial	Elaboración automática de gráficos, tablas e informes financieros o presupuestarios basados en datos de acceso público o no confidenciales, garantizando la exclusión de información personal o sensible.	Todas las Subdirecciones funcionales (varía por materia)	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Asistencia en la aproximación y preparación general de argumentos, estudios jurídicos y técnicos	Utilización de sistemas de IA generativa para el análisis y contraste de fuentes jurídicas, doctrinales o técnicas de carácter público o no confidencial, con el fin de elaborar marcos argumentales, estudios comparados, notas técnicas o hipótesis interpretativas. Estas herramientas apoyan la preparación de estrategias o enfoques jurídicos sin sustituir la valoración profesional ni implicar toma de decisiones automatizadas. Asimismo, pueden emplearse para el análisis exploratorio y sistemático de fuentes jurídicas o abiertas, la comparación de marcos normativos o doctrinales, y la elaboración de estudios generales que no involucren información confidencial ni datos personales.	Todas las Subdirecciones funcionales (varía por materia)	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Asistencia en la redacción inicial de elementos generales y aislados cuyas ideas luego puedan incorporarse a resoluciones, informes o guías técnicas	Utilización de sistemas de IA generativa para el análisis y contraste de fuentes jurídicas, doctrinales o técnicas de carácter público o no confidencial, con el fin de elaborar marcos argumentales, estudios comparados, notas técnicas o hipótesis interpretativas. Estas herramientas apoyan la preparación de estrategias o enfoques jurídicos sin sustituir la valoración profesional ni implicar toma de decisiones automatizadas. Asimismo, pueden emplearse para el análisis exploratorio y sistemático de fuentes jurídicas o abiertas, la comparación de marcos normativos o doctrinales, y la elaboración de estudios generales que no involucren información confidencial ni datos personales.	Todas las Subdirecciones funcionales (varía por materia)	Bajo	Sistema Externo	Revisión editorial Sin datos personales, ni información sensible/confidencial.
Asistencia en tareas de desarrollo y configuración de sistemas	La IA se ha convertido en una herramienta clave para tareas técnicas como scripting, generación de consultas SQL, construcción de expresiones regulares y resolución de dudas sobre software específico. Facilita gran parte del trabajo, permitiendo a los usuarios enfocarse	Secretaría General	Bajo-Medio	Sistema Externo / Interno	Revisión editorial Sin datos personales, ni información sensible/confidencial, ni información sobre arquitecturas y configuraciones de los sistemas corporativos.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
	en la revisión y ajuste fino en lugar de empezar desde cero. Esto es especialmente útil en entornos donde se manejan múltiples tecnologías y configuraciones avanzadas de sistemas.				
Asistente interno normativo o doctrinal	Desarrollo de asistentes virtuales internos que faciliten la consulta rápida de legislación, resoluciones, informes jurídicos, criterios interpretativos de la Agencia, etc., mejorando el acceso al conocimiento organizativo. Pueden servir de apoyo para la inspección, para las distintas fases de tramitación de solicitudes de aprobación de códigos de conducta y las de solicitudes de acuerdos de transferencias internacionales, o para la elaboración de resúmenes anonimizados para apoyo a divulgación de contenidos, entre otras.	Todas las Subdirecciones funcionales (varía por materia)	Bajo (si únicamente se usan documentos públicos) -Medio	Sistema Externo (si únicamente se usan documentos públicos) Sistema Interno / Sistema Ad-hoc	Revisión editorial. Sin datos personales, ni información sensible/confidencial. No debe generar contenido vinculante. Actualización corpus documental periódica.
		Subdirección General de Promoción y autorizaciones	Medio	Sistema Interno / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
		Prensa y comunicaciones	Medio	Sistema Interno / Sistema Ad-hoc	Revisión editorial. Sin datos personales, ni información sensible/confidencial.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
					No debe generar contenido vinculante. Actualización corpus documental periódica.
		Subdirección General de Inspección	Medio-alto si utiliza datos personales o información sensible/confidencial, o implica decisiones que pueden impactar en terceros.	Sistema Interno / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
Transcripción de audio y vídeo procedente de fuentes abiertas sin información confidencial	Transcripción de contenidos multimedia de fuentes abiertas o no confidenciales para su incorporación a diferentes flujos de trabajo o estudios. Podría contener datos personales.	Todas las Subdirecciones funcionales (varía por materia)	Medio	Sistema Interno	Revisión editorial. Sin información sensible/confidencial. No debe generar contenido vinculante.
Transcripción de audio y vídeo que contenga información interna, privada o confidencial, en su caso de reuniones y apoyo a la generación de actas	Transcripción de reuniones internas, en su caso para apoyo a la generación de actas. Así como de audios particulares de entrevistas concedidas o intervenciones realizadas por el personal de la Agencia en actos o eventos y elaboración de resumen del texto resultante.	Todas las Subdirecciones funcionales (varía por materia) Prensa y comunicación	Medio	Sistema Interno	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Estructuración, gestión o resúmenes de documentos administrativos internos o con datos personales	Aplicación de modelos inteligentes para resumir, extraer información clave y convertir documentos administrativos internos en estructuras normalizadas que mejoren su tramitación o análisis automatizado. En estos casos, debe aplicarse una anonimización completa o parcial de los datos personales y limitar el tratamiento a entornos seguros y autorizados, por ejemplo, para elaborar versiones anonimizadas destinadas a consultas de transparencia o informes internos.	Todas las Subdirecciones funcionales (varía por materia)	Medio Datos personales o confidenciales/sensibles.	Sistema Interno / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
Redacción de borradores, cartas, correos o notas internas con tono institucional	Utilización de sistemas inteligentes para elaborar propuestas iniciales de respuesta a ciudadanos u organismos, o escritos agilizando tiempos de tramitación y garantizando coherencia en los mensajes institucionales.	Secretaría General	Medio	Sistema Interno / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
Generación de borradores de respuesta a consultas para canales de atención al ciudadano,	La utilización de RAGs y sistemas bien afinados permitirán acceder a información actualizada en tiempo real sobre todo el material producido por la AEPD para producir borradores de respuestas a consultas específicas de los canales de atención al	Subdirección General de Promoción y autorizaciones	Medio-alto	Sistema interno/ Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
canal DPD y joven.	ciudadano, canal del DPD y canal joven.				
Clasificación y resumen de denuncias, reclamaciones, consultas y otras entradas.	Aplicación de modelos de procesamiento de lenguaje natural (PLN) para facilitar el tratamiento inicial de la información que llega a la Agencia, permitiendo su categorización, indexación y análisis preliminar.	Subdirección General de promoción y autorizaciones	Medio	Sistema Interno/Sistema Ad-Hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
Apoyo en el análisis del EIPD (Evaluación de Impacto en Protección de Datos)	Utilización de modelos expertos para realizar un primer triaje o apoyo documental en los informes de evaluación de impacto, proporcionando plantillas, criterios comunes o guías de análisis.	Subdirección General de Inspección de Datos/División de Innovación Tecnológica/DPD	Medio-alto	Sistema Interna / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
Gestión inteligente del plan estratégico	La automatización de la gestión del plan estratégico facilita el seguimiento en tiempo real de los indicadores relativos al cumplimiento de los objetivos y resultados. Permite automatizar la metodología escogida, optimiza la ejecución de la estrategia y permite la toma de decisiones basada en datos en tiempo real. Análisis de datos y	Presidencia/Adjuntía	Alto	Sistema Interno / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
	visualización de resultados para seguimiento de planes e indicadores.				
Asistencia en la redacción de resoluciones, informes o guías técnicas.	Uso de herramientas de IA generativa como apoyo a la elaboración inicial de borradores de documentos normativos y doctrinales, facilitando la estructuración de contenidos y aumentando la eficiencia en la producción documental.	Subdirección General de Inspección de Datos/División de Innovación Tecnológica/DPD	Medio-alto	Sistema Interno / Sistema Ad-hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.
Sistemas de alertas inteligentes (priorización de denuncias, reclamaciones, notificaciones y otras entradas)	Desarrollo de sistemas de IA o RPA que permitan detectar y priorizar denuncias, brechas de datos o comunicaciones sensibles, incluyendo alertas automáticas para casos de alto impacto, colectivos vulnerables o situaciones que requieran una atención prioritaria a criterio de la Agencia, permitiendo así una gestión ágil y focalizada.	Subdirección General de Inspección de Datos/División de Innovación Tecnológica/ Gabinete de Prensa y Comunicación	Alto	Sistema Interno / Sistema Ad-Hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.

Caso de uso	Descripción	División / Subdirección principal	Impacto institucional o funcional estimado	Tipo de sistema recomendado	Observaciones / Obligaciones específicas
Soporte a la tramitación de expedientes y notificaciones de brechas de datos	La tramitación de expedientes y/o notificaciones de brechas de datos personales son procesos manuales asistidos por varias herramientas corporativas. La IAG permite acelerar tareas, desde la clasificación inicial de las entradas por registro, su categorización y extracción de información estructurada, hasta la elaboración de resúmenes y borradores.	Subdirección General de Inspección de Datos/División de Innovación Tecnológica	Alto	Sistema Interno/ Sistema Ad-Hoc	Revisión editorial y control humano si implica decisiones. No debe generar contenido vinculante. Actualización corpus documental periódica.

B. ANÁLISIS DE LAS AMENAZAS EN DETALLE CON RELACIÓN AL SISTEMA DE IAG IMPLEMENTADO EN LOS PROCESOS

A continuación, se presentan ejemplos representativos de amenazas identificadas en la aplicación de sistemas de inteligencia artificial generativa (IAG) a procesos administrativos, basadas en los casos de uso analizados en la AEPD. Estas amenazas no constituyen un análisis exhaustivo, sino una referencia inicial que se adaptará o extenderá según sea necesario. Están amenazas se gradúan en función del tipo de sistema IAG que se implemente en los procesos:

- Sistema Externo (Sistemas IAG de terceros desplegados en infraestructura fuera del control de la organización, utilizados como SaaS bajo sus términos de uso)
- Sistema Interno (Sistemas IAG desarrollados por terceros y desplegados en infraestructura bajo control de la organización)
- Sistema Ad-hoc (Sistemas IAG desarrollados internamente o por terceros bajo especificaciones a medida y desplegados en infraestructura bajo control de la organización e integrado con sistemas internos

El propósito de la política general planteada por la AEPD es gestionar adecuadamente dichas amenazas para, bien reducir su impacto, o bien reducir la probabilidad de su materialización, mediante la oportuna selección de sistemas de IAG por casos de uso, la aplicación de medidas organizativas y técnicas, así como la ejecución de los mecanismos de supervisión adecuados de la gobernanza y la gestión efectiva de las políticas.

En las siguientes tablas se describen ejemplos de amenazas técnicas y de funcionamiento que pueden afectar a los sistemas de inteligencia artificial generativa (IAG), clasificadas según el tipo de sistema (externo, interno o ad-hoc). Las valoraciones Alto / Medio / Bajo no representan una probabilidad ni un impacto cuantitativo, sino una estimación cualitativa del nivel de exposición o vulnerabilidad del sistema frente a cada amenaza, en función del grado de control que la organización puede ejercer sobre el modelo, los datos y el entorno operativo.

1. Amenazas a la eficacia de los sistemas de IAG

La eficacia de los sistemas de IA afecta, entre otros, a la protección de derechos fundamentales en la medida que la competencia de la AEPD es, precisamente, la protección de los mismos con relación al tratamiento de datos personales.

Amenaza	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Alucinaciones: Respuestas plausibles pero falsas o inventadas	**Alto**: sin control sobre el modelo, difícil detectar errores sistemáticos.	**Medio**: posibilidad de incorporar mecanismos de revisión.	**Bajo**: adaptado y/o entrenado con corpus controlado y diseñado para el dominio.
Falta de conocimiento específico: Incapacidad del modelo para responder con precisión en ámbitos especializados.	**Alto**: modelos generalistas, entrenados para amplios contextos.	**Medio**: se pueden conectar a fuentes internas.	**Bajo**: adaptado y/o entrenado para el caso de uso con datos del entorno AEPD.

Correlaciones irrelevantes	**Medio**: modelo opaco, sin posibilidad de ajustar comportamiento.	**Medio**: control parcial, se pueden ajustar parámetros de entrada/salida.	**Bajo**: adaptación y/o diseño de entrenamiento más riguroso, menor riesgo de errores lógicos.
Salidas no repetibles: Variabilidad en las respuestas ante entradas iguales, dificultando la coherencia, debido a que no existe un control del contexto al que hace acceso el modelo	**Alto**: comportamiento impredecible en cada llamada a la API.	**Bajo**: configuración/parametrización del modelo bajo control.	**Bajo**: configuración/parametrización del modelo bajo control, posibilidad de estandarizar salidas.
Obsolescencia tecnológica acelerada, especialmente si se adoptan soluciones poco maduras o con fuerte dependencia del proveedor	**Alto**: soluciones de mercado pueden quedar desactualizadas o abandonadas.	**Medio**: depende del ciclo de vida del modelo y de la capacidad técnica interna.	**Bajo**: diseño adaptado al contexto institucional con posibilidad de evolución planificada.

2. Amenazas de sesgo y discriminación

Amenaza	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Sesgos en datos de entrenamiento: Reducción del pensamiento crítico y validación humana.	**Alto**: no se conoce el origen ni se puede auditar. El modelo puede arrastrar sesgos culturales o demográficos.	**Medio**: posibilidad de revisar parte del modelo o aplicar filtros, aunque el dataset original no siempre es accesible.	**Bajo**: se puede seleccionar, curar y auditar el dataset para asegurar diversidad, equilibrio y representatividad.
Sesgos algorítmicos	**Alto**: no es posible modificar la arquitectura ni conocer en detalle los sesgos algorítmicos incorporados.	**Medio**: se pueden aplicar ajustes menores (reinstrucción, afinado) pero no se accede al diseño del modelo base.	**Bajo**: el diseño del modelo, la selección de hiperparámetros y el entrenamiento están en manos del organismo. Se

			puede reducir la introducción de sesgos desde el inicio.

3. Impactos para los derechos y libertades con relación a la protección de datos

Este conjunto de amenazas aplica fundamentalmente cuando el sistema IAG se emplea en procesos de la AEPD que implican datos personales, como podrían ser los que se derivan de casos de uso en los que se tratan textos que contienen información de personas físicas.

Existen multiplicidad de casos de uso en los que los casos de uso no conllevarían explícitamente el tratamiento de datos de personas físicas, como es el caso de análisis o resúmenes de normas. En esos casos, y con un menor impacto en la mayor parte de ellos, hay que tener en cuenta las amenaza que suponen la recogida de metadatos y perfilado de los propios usuarios de los sistemas IAG, o bien la manipulación y robo de información de los propios usuarios mediante técnicas de inyección indirecta de instrucciones (prompts).

Incidente o escenario de riesgo	Amenazas según metodología LIINE4DU[1]	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Fugas o uso indebido de datos personales: de otros miembros de la organización o de terceros.	Vinculación Identificación Brecha de datos Divulgación	**Muy alto**: riesgo de reutilización para entrenamiento, falta de garantías contractuales, registros en servidores del proveedor.	**Bajo**: el entorno es controlado por la AEPD; se pueden aplicar medidas ENS y políticas internas.	**Muy bajo**: todo el ciclo de tratamiento está bajo supervisión institucional, sin transferencia externa.
Exposición de datos personales durante el entrenamiento o uso	Vinculación Identificación Brecha de datos Divulgación	**Alto**: los datos enviados a través de prompts pueden quedar registrados o ser reutilizados por el proveedor.	**Bajo**: si los datos se usan en un entorno cerrado sin conexión externa.	**Muy bajo**: los datos se seleccionan y gestionan de forma segura; posibilidad de anonimización previa.

[1] Introducción a liine4du 1.0: una nueva metodología para el modelado de amenazas para la privacidad y la protección de datos

Perfilado del usuario (persona física) de la IAG a través de la recogida de prompts, metadatos o acceso a información personal almacenada en el sistema	Vinculación Identificación Detección Engaño	**Muy alto**: los términos de uso de proveedores externos pueden incluir reentrenamiento o análisis de entradas.	**Bajo**: en entorno local, el tratamiento está limitado por las propias políticas técnicas.	**Inexistente**: los datos no abandonan el entorno institucional, ni se comparten ni se reutilizan externamente.
Dificultad para borrar datos relativos a personas físicas procesados en la interacción con el servicio.	No repudio Desconocimiento y falta de capacidad para intervenir	**Muy alto**: no hay garantías de supresión efectiva ni control sobre backups o modelos reentrenados.	**Medio**: depende del almacenamiento y configuración del modelo local.	**Bajo**: posibilidad de usar técnicas reversibles (LoRA, Adapters) y aplicar borrado controlado.
Accesos no autorizados a documentos internos o fuentes con datos personales	Brecha de datos Divulgación	**Alto**: mayor superficie de ataque y menor control sobre accesos y almacenamiento.	**Medio**: riesgo mitigado mediante medidas técnicas (ENS, roles, cifrado).	**Bajo**: control completo del entorno y políticas de acceso.
Falta de trazabilidad y auditoría de los datos utilizados	Inexactitud Desconocimiento y falta de capacidad para intervenir	**Muy alto**: opacidad total del modelo y de los procesos internos del proveedor.	**Medio**: trazabilidad parcial posible si se documentan procesos locales.	**Bajo**: trazabilidad garantizada si se define desde el diseño.
Persistencia de datos personales en logs en manos del proveedor del servicio o tercero	Vinculación Identificación Brecha de datos Divulgación	**Alto**: los registros pueden mantenerse fuera del control de la AEPD.	**Medio**: requiere control interno de registros y eliminación periódica.	**Bajo**: se pueden anonimizar o purgar automáticamente.
Pérdida de control sobre modelos preentrenados	Inexactitud Exclusión	**Muy alto**: imposible conocer los datos usados ni modificar el modelo.	**Medio**: menor riesgo, aunque sigue sin poder auditar el preentrenamiento.	**Bajo**: modelo completamente trazable y entrenado con datos conocidos.
Vulnerabilidades en interfaces, APIs o ataques adversarios o puertas traseras	Vinculación Identificación Brecha de datos Divulgación	**Alto**: alta exposición en entornos en línea, APIs abiertas o no auditadas.	**Medio**: mitigable con buenas prácticas de ciberseguridad.	**Bajo**: menor superficie de ataque y diseño cerrado según ENS.
Seguridad física de personas protegidas si se expone información	Vinculación Identificación Brecha de datos	**Bajo**: existiría riesgo si se usan datos operativos o agendas sensibles en sistemas	**Bajo**: sería mitigable si se segmentan accesos y se	**Bajo**: control completo del entorno

sobre agendas, movimientos o decisiones sensibles	Divulgación	accesibles, que no es el caso de la AEPD.	blindan los datos, pero no es al caso de la AEPD.	y exclusión de información sensible.

4. Amenazas a la seguridad de la infraestructura y continuidad de procesos de negocio

Incidente o escenario de riesgo	Amenazas según STRIDE[2]	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Vulnerabilidades técnicas en los sistemas que alojan los sistemas IAG	Suplantación de identidad *Tampering* Denegación de servicio	**Alto**: expuesto a vectores externos, sin control total sobre parches o capas de seguridad.	**Medio**: mitigable con medidas ENS, segmentación y hardening.	**Bajo**: entorno cerrado, diseñado y asegurado según estándares internos.
Acceso a internet no controlado, con riesgo de amenazas externas.	*Tampering* Revelación de información Escalado de privilegios	**Alto**: riesgo de entrada de malware, conexión a fuentes inseguras, descarga de contenidos maliciosos.	**Medio**: si se restringe el acceso y se monitoriza adecuadamente.	**Bajo**: se puede aislar totalmente si es necesario.
Inyección de prompts o ataques adversarios	*Tampering* Revelación de información Escalado de privilegios	**Alto**: proveedores externos son más susceptibles a ataques sofisticados sin supervisión directa.	**Medio**: mitigable con validación de entradas y control de interfaces.	**Bajo**: permite protección desde el diseño y control total del input.
Errores humanos en la gestión técnica o configuración del entorno.	Suplantación de identidad *Tampering* Repudiación Revelación de información Denegación de servicio Escalado de privilegios	**Alto**: dependencia de terceros, desconocimiento del backend.	**Medio**: gestionable con formación y procedimientos internos.	**Bajo**: control completo por personal especializado y formación específica.

[2] OWAST Threat Modeling Process

Resiliencia y disponibilidad: Riesgo de interrupción del servicio por falta de soporte, licencias o discontinuidad del producto	Denegación de servicio	**Alto**: riesgo de cambios unilaterales de condiciones, cancelación de servicios, falta de soporte o licencias.	**Medio**: dependerá del mantenimiento técnico y soporte interno.	**Bajo**: la continuidad depende del plan interno de evolución y escalabilidad.
Portabilidad del sistema: Dificultad para migrar o sustituir el sistema por falta de interoperabilidad o dependencia tecnológica	Denegación de servicio	**Alto**: bajo control sobre el formato, código y modelo. Dificultad de migración sin pérdida.	**Medio**: mayor flexibilidad si se gestiona el ciclo técnico completo.	**Bajo**: arquitectura diseñada para ser reutilizable o migrable.
Posible exposición de información sensible vinculada a personas físicas o altos cargos institucionales	Suplantación de identidad Revelación de información	**Alto**: riesgo grave si se integra en entornos abiertos o sistemas no controlados.	**Medio**: se puede proteger con cifrado, roles y segmentación.	**Bajo**: solo accesible desde canales seguros definidos por la AEPD.

5. Divulgación de información no personal (sensible/confidencial de las actuaciones de la AEPD)

Amenaza	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Divulgación institucional al exponer criterios internos, estrategias o enfoques que puedan afectar la imagen o la autoridad de la AEPD	**Muy alto**: riesgo de memorizar prompts o respuestas con información crítica institucional sin garantías de borrado o confidencialidad.	**Medio**: control limitado si se reutilizan modelos cerrados.	**Bajo**: contenido y uso totalmente supervisados; sin riesgo de exposición fuera del entorno.
Revelación de intereses estratégicos al dejar entrever áreas prioritarias de actuación que podrían ser utilizadas por terceros para anticiparse.	**Alto**: los contenidos generados pueden exponer indirectamente líneas prioritarias o temas sensibles.	**Medio**: riesgo si no se controlan inputs/outputs y corpus indexado.	**Bajo**: posibilidad de definir reglas de exclusión de temas sensibles.
Revelación de actuaciones de terceros sometidos a inspección mediante contenidos generados por IA que permitan deducir o identificar procedimientos aún no públicos	**Muy alto**: el modelo puede devolver información inferida que vincule indirectamente a actuaciones reales pasadas o en curso.	**Medio**: mitigable con control de fuentes y prompts.	**Bajo**: supervisión del contenido y de las salidas generadas.
Exposición de documentos de trabajo internos como borradores, actas o	**Alto**: riesgo elevado si se utilizan borradores, actas o textos	**Medio**: se puede evitar con segmentación de acceso y control documental.	**Bajo**: los documentos se mantienen en repositorios

intercambios que comprometan la confidencialidad organizativa	confidenciales en prompts o como fuentes en sistemas RAG externos.		internos con gestión de acceso controlada.
Intoxicación de las fuentes de datos para manipular respuestas, acciones y robo de información corporativa (ataques de prompting indirecto)	**Alto**: si el sistema RAG accede a fuentes en línea, podría incorporar datos manipulados o falsos.	**Medio**: puede ocurrir si no se filtra bien el corpus interno.	**Bajo**: corpus curado, validado y gestionado por el organismo. Control sobre el ciclo de ingestión.
Filtración de relaciones con otras autoridades o entidades públicas/privadas, afectando la cooperación institucional o la neutralidad percibida	**Alto**: posibilidad de revelar colaboraciones, intercambios o actuaciones conjuntas no publicadas.	**Medio**: depende del corpus utilizado y de los logs generados.	**Bajo**: la adaptación y/o el entrenamiento y uso están bajo estricta supervisión funcional y legal.

6. Divulgación de información no personal (sensible/confidencial de terceros)

Amenaza	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Divulgación de secretos comerciales aportados en procedimientos o consultas, al ser utilizados como ejemplos o prompts sin protección	**Muy alto**: posibilidad de retención y uso no controlado de ejemplos o contenidos sensibles en los prompts.	**Medio**: riesgo si se usan sin etiquetado o aislamiento en el corpus.	**Bajo**: adaptación y/o entrenamiento controlado con exclusión explícita de información clasificada.
Exposición de información técnica o jurídica confidencial incluida en documentación presentada por empresas, organismos o asesores	**Alto**: modelos externos podrían registrar o inferir patrones de contenido sensible.	**Medio**: mitigable si se aplican filtros y restricciones documentales.	**Bajo**: posibilidad de excluir categorías o marcar documentos confidenciales en el diseño del sistema.
Filtración de elementos protegidos por propiedad intelectual como software, algoritmos o modelos, si se reutilizan como parte del entrenamiento	**Muy alto**: riesgos legales si se incorpora software, lógica de negocio o documentación técnica a modelos sin control.	**Medio**: requiere control estricto del origen de datos y licencias.	**Bajo**: adaptación y/o entrenamiento con materiales autorizados o generados por la AEPD.
Pérdida de confianza institucional por parte de terceros al percibir riesgo de reutilización o exposición de la información facilitada	**Alto**: uso de plataformas comerciales puede generar desconfianza sobre la protección de la información aportada.	**Medio**: si no se garantiza la trazabilidad y protección documental.	**Bajo**: control del entorno y documentación de garantías ofrecidas a terceros.

Revelación no autorizada de contenidos marcados como confidenciales por otros organismos, empresas o entidades del sector público	**Alto**: posible reutilización o inferencia de contenidos marcados como confidenciales por otros organismos.	**Medio**: mitigable con políticas internas de clasificación y segmentación.	**Bajo**: cumplimiento estricto de cláusulas y control del ciclo de datos.

7. Interacción humana incorrecta, irresponsable o perjudicial con la IAG

Amenaza	Sistema Externo	Sistema Interno	Sistema Ad-hoc
Incumplimiento de las políticas de uso de sistemas IA con relación al uso de información personal, confidencial o sensible.	**Alto:** No hay medidas por diseño que impidan	**Bajo:** imposible por medidas desde el diseño	**Bajo:** imposible por medidas desde el diseño
Impacto negativo sobre las condiciones laborales del personal, en caso de que la implantación IAG en procesos suponga una cosificación de los empleados o se presione para un aumento de la productividad más allá de un uso racional de la IAG	**Medio:** no dependiente del sistema		
Resistencia al cambio	**Medio:** no dependiente del sistema		
Percepción de la IAG como una amenaza: como monitorización excesiva o indebida del rendimiento laboral mediante sistemas IAG.	**Medio:** no dependiente del sistema		
Uso incorrecto de las herramientas con pérdida de efectividad	**Medio:** no dependiente del sistema		
Interacción poco crítica con los sistemas inteligentes puede ser fuente de decisiones automatizadas, errores, sesgos o decisiones inapropiadas	**Alto:** no dependiente del sistema		

8. Falta de transparencia y explicabilidad de las actuaciones basadas en IAG o falta de coherencia ante situaciones similares o desviaciones en la aplicación de criterios vigentes

Amenaza	Sistema Externo	Sistema Interno	Sistema Ad-hoc

Percepción de cosificación y deshumanización en procesos orientados al ciudadano	**Medio:** No depende del sistema		
Reproducción o amplificación de desigualdades sociales si el sistema discrimina indirectamente a colectivos vulnerables	**Alto**: el entrenamiento externo puede contener sesgos no corregidos.	**Medio**: control parcial si se revisa el dataset o se aplica fine-tuning.	**Medio**: posible eliminar sesgos desde la adaptación y/o selección del corpus y los criterios de entrenamiento.
Pérdida de confianza ciudadana en la AEPD si se percibe que las decisiones o actuaciones están influenciadas por sistemas automáticos opacos	**Muy alto**: el uso de plataformas comerciales opacas puede erosionar la percepción de imparcialidad.	**Medio**: si no se comunica adecuadamente el alcance del uso de IA.	**Bajo**: alta trazabilidad, transparencia y gobernanza visible.
Impacto social de una brecha	**Alto:** al existir una comunicación de datos y dependencia de sistemas externos.	**Bajo:** se minimiza el proceso externo.	**Bajo:** se minimiza el proceso externo.
Falta de explicabilidad	**Alto:** generalmente, no se ofrece información de explicabilidad y las pruebas realizadas por el usuario no controlan todos los parámetros.	**Medio:** generalmente, no se ofrece información de explicabilidad pero hay un mayor control de pruebas.	**Bajo:** Se puede controlar y obtener información de explicabilidad.
Confianza excesiva del usuario	**Alto**: la opacidad del sistema externo y su apariencia de autoridad aumentan el riesgo de aceptación acrítica.	**Medio**: más control sobre el entorno puede favorecer un uso más consciente, pero aún puede haber exceso de confianza.	**Bajo**: al formar parte de un proceso institucional controlado, es más fácil implantar formación, validación y revisión sistemática.
Fallos en la operación del sistema IA	**Alto:** hay una menor posibilidad de control de la eficacia de los sistemas	**Medio:** hay una mayor posibilidad de control de la eficacia de los sistemas	**Medio:** hay una mayor posibilidad de control de la eficacia de los sistemas

9. Desgobierno y pérdida de integridad institucional

Amenazas	Sistema Externo	Sistema Interno	Sistema Ad-hoc

Fraude o manipulación de procedimientos	**Alto**: mayor exposición si no se conocen las lógicas internas ni se puede auditar el comportamiento del sistema.	**Medio**: mitigable con validaciones cruzadas y logs locales.	**Bajo**: diseño bajo principios de control interno y trazabilidad.
Impacto financiero interno por mantenimiento, licencias, escalabilidad o sobrecostes inesperados.	**Alto**: licencias, precios variables, lock-in tecnológico y baja previsión a largo plazo.	**Medio**: costes controlables, puede haber dependencia de actualizaciones o soporte externo.	**Medio**: mayor inversión inicial, pero sostenibilidad económica planificable.
Pérdida de control institucional sobre funciones críticas si se delegan en exceso en tecnologías sin reversibilidad o capacidad de auditoría	**Muy alto**: si decisiones relevantes quedan delegadas en modelos opacos o de terceros.	**Medio**: mitigable con validación y trazabilidad.	**Bajo**: decisiones permanecen en el ámbito institucional.
Disfunciones en la coordinación interadministrativa si no se siguen estándares comunes o se introducen sistemas sin interoperabilidad	**Alto**: riesgo si se usan soluciones no interoperables o no alineadas con estándares AGE.	**Medio**: se puede alinear con políticas tecnológicas comunes.	**Bajo**: diseño conforme a criterios de interoperabilidad y estándares públicos.

C. MEJORA DE RESULTADOS Y ADAPTACIÓN AL DOMINIO DE USO

Para mejorar el desempeño y la adaptación de los modelos a necesidades específicas se usan dos técnicas, que pueden complementarse con técnicas de aprendizaje reforzado, como son:

- Fine-tuning o ajuste fino o específico: Partiendo de un modelo preentrenado, se reajustan sus parámetros (algunos o todos) mediante un reentrenamiento sobre un conjunto de datos específicos. Este proceso permite adaptar un LLM a una terminología, estilo y requerimientos específicos.
- Retrieval-Augmented Generation (RAG) o generación aumentada por recuperación: No se modifica el modelo, si no que existe una fase previa de recuperación de información en tiempo real desde bases de datos, documentos, internet u otras fuentes (*retrieval*). Esta información recuperada se añade al *prompt* de entrada del LLM. Esto permite acceder a información actualizada sin necesidad de reentrenar el modelo.[3]

D. PLAN DE DESPLIEGUE

Un plan estructurado para alcanzar los objetivos del modelo de gobernanza tendrá que incluir los siguientes hitos (que pueden solaparse):

- Definición de objetivos y alcance: Establecer los objetivos específicos del modelo de gobernanza y determinar su alcance, identificando las áreas y procesos donde la IAG se puede generar mayor valor y que serán cubiertos. Involucrar al personal y a los usuarios en la identificación de necesidades (de 1 a 2 meses).
- Evaluación de la situación actual: Identificar y catalogar los datos disponibles, los procedimientos y procesos existentes, la infraestructura tecnológica, detectando problemas potenciales. Evaluar la viabilidad y el impacto de la solución de IAG (2 meses).
- Diseño del modelo de gobernanza: Definir roles y responsabilidades, establecer las políticas y procedimientos correspondientes (3 meses).
- Desarrollo de capacidades y formación: Capacitar al personal implicado y sensibilizar al resto del personal. Organización de cursos y talleres de formación, material de apoyo y guías prácticas (de 4 a 6 meses).
- Implementación del modelo de gobernanza e infraestructura: Poner en marcha las políticas y procedimientos definidos. Implementar la infraestructura técnica. Formalizar contratos con proveedores y encargados. Integrar medidas de supervisión y auditoría (de 4 a 6 meses).
- Fase de pruebas y piloto: implementar prototipos de la solución de IAG seleccionada, evaluando y refinando su desempeño primero en un entorno de

[3] Ver por ejemplo la guía publicada por el CCN para crear un Chatbot que usa modelos de terceros de forma local con opción de incluir también 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).

pruebas y luego un piloto en un entorno real, en una selección de los casos de uso posibles. Ajustar y optimizar el sistema (de 3 a 6 meses, más si se usa un modelo ad-hoc).

- Implementar el sistema de IAG en todos los casos de uso (de 3 a 6 meses o más, según el número de casos de uso y su dimensión).
- Monitorización y mejora continua: Evaluación continua del modelo de gobernanza. Actualización de políticas y procedimientos según sea necesario. Mantener registros detallados de las actividades. Expansión de los casos de uso y extrapolación a otros procesos. Realización de las auditorías (cada 3 o 4 meses, al menos anualmente).

A continuación, se muestra un ejemplo de planificación provisional: