centro criptológico nacional

Guía de Seguridad de las TIC
CCN-STIC 863

Uso seguro de los sistemas de inteligencia artificial en el marco del ENS



Noviembre 2026

Edita:

centro criptológico nacional

© Centro Criptológico Nacional, 2026

centro criptológico nacional

Firmado digitalmente por CCN
Fecha: 2026.11.26 21:54:24 +01'00'

Fecha de Edición: noviembre de 2026

LIMITACIÓN DE RESPONSABILIDAD

El presente documento se proporciona de acuerdo con los términos en él recogidos, rechazando expresamente cualquier tipo de garantía implícita que se pueda encontrar relacionada. En ningún caso, el Centro Criptológico Nacional puede ser considerado responsable del daño directo, indirecto, fortuito o extraordinario derivado de la utilización de la información y software que se indican incluso cuando se advierta de tal posibilidad.

AVISO LEGAL

Quedan rigurosamente prohibidas, sin la autorización escrita del Centro Criptológico Nacional, bajo las sanciones establecidas en las leyes, la reproducción parcial o total de este documento por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares del mismo mediante alquiler o préstamo públicos.

ÍNDICE

1. INTRODUCCIÓN 5
1.1. OBJETO 5
1.2. ALCANCE 6
1.3. RELACIÓN CON EL ENS Y CON OTRAS GUÍAS DEL CCN 6
2. CONCEPTOS Y MODELO DE REFERENCIA 7
3. INVENTARIO DE ACTIVOS 10
4. CLASIFICACIÓN Y GESTIÓN DE RIESGOS EN SISTEMAS DE IA 11
4.1. CLASIFICACIÓN DE SISTEMAS DE IA CONFORME AL RIA 11
4.2. RIESGOS ASOCIADOS AL USO DE SISTEMAS DE IA 13
5. MARCO DE GOBIERNO 17
5.1. PRINCIPIOS GENERALES 17
5.2. INTEGRACIÓN EN EL MARCO DE GOBIERNO 18
5.2.1. GOBIERNO Y DIRECCIÓN 18
5.2.2. SUPERVISIÓN 19
5.2.3. OPERACIÓN 21
5.3. FORMACIÓN, CONCIENCIACIÓN Y ALFABETIZACIÓN EN IA 21
6. IMPLANTACIÓN DE LA SEGURIDAD 22
6.1. CRITERIOS DE APLICABILIDAD DE LAS MEDIDAS DEL ENS 22
6.1.1. [ORG.1] POLÍTICA DE SEGURIDAD 22
6.1.2. [ORG.2] NORMATIVA DE SEGURIDAD 22
6.1.3. [ORG.3] PROCEDIMIENTOS DE SEGURIDAD 23
6.1.4. [ORG.4] PROCESO DE AUTORIZACIÓN 23
6.1.5. [OP.PL.1] ANÁLISIS DE RIESGOS 23
6.1.6. [OP.PL.2] ARQUITECTURA DE SEGURIDAD 23
6.1.7. [OP.PL.3] ADQUISICIÓN DE NUEVOS COMPONENTES 23
6.1.8. [OP.PL.4] DIMENSIONAMIENTO / GESTIÓN DE LA CAPACIDAD 24
6.1.9. [OP.ACC.1] IDENTIFICACIÓN 24
6.1.10. [OP.ACC.2] REQUISITOS DE ACCESO 24
6.1.11. [OP.ACC.3] SEGREGACIÓN DE FUNCIONES Y TAREAS 24
6.1.12. [OP.ACC.4] PROCESO DE GESTIÓN DE DERECHOS DE ACCESO 24
6.1.13. [OP.EXP.1] INVENTARIO DE ACTIVOS 25
6.1.14. [OP.EXP.2] CONFIGURACIÓN DE SEGURIDAD 25
6.1.15. [OP.EXP.3] GESTIÓN DE LA CONFIGURACIÓN DE SEGURIDAD 25
6.1.16. [OP.EXP.5] GESTIÓN DE CAMBIOS 26
6.1.17. [OP.EXP.7] GESTIÓN DE INCIDENTES 26
6.1.18. [OP.EXP.8] REGISTRO DE LA ACTIVIDAD 26
6.1.19. [OP.NUB.1] PROTECCIÓN DE SERVICIOS EN LA NUBE 26
6.1.20. [OP.EXT.1] CONTRATACIÓN Y ACUERDOS DE NIVEL DE SERVICIO 27
6.1.21. [OP.EXT.3] PROTECCIÓN DE LA CADENA DE SUMINISTRO 27
6.1.22. [OP.EXT.4] INTERCONEXIÓN DE SISTEMAS 27
6.1.23. [OP.MON.1] DETECCIÓN DE INTRUSIÓN 27
6.1.24. [OP.MON.3] VIGILANCIA 28
6.1.25. [MP.PER.3] CONCIENCIACIÓN 28

6.1.26. [MP.PER.4] FORMACIÓN 28
6.1.27. [MP.SW.1] DESARROLLO DE APLICACIONES 29
6.1.28. [MP.INFO.1] DATOS PERSONALES 29
6.1.29. [MP.INFO.6] COPIA DE SEGURIDAD 29
6.1.30. [MP.S.2] PROTECCIÓN DE SERVICIOS Y APLICACIONES WEB 29
6.2. MATRIZ DE ESCENARIOS Y MEDIDAS RELACIONADAS 30
7. GLOSARIO DE TÉRMINOS 32
8. MARCO NORMATIVO Y DE REFERENCIA 36
9. ACRÓNIMOS 37
ANEXO A CASO DE USO: ASISTENTE NORMATIVO 38
A1. DESCRIPCIÓN DEL CASO DE USO 38
A2. ARQUITECTURA DEL SISTEMA DE IA 38
A3. INVENTARIO DE ACTIVOS 40
A4. TIPO DE SISTEMA DE IA 41
A5. RIESGOS ESPECÍFICOS IDENTIFICADOS 41
A6. APLICACIÓN DE MEDIDAS DE SEGURIDAD 42
ANEXO B CASO DE USO: SISTEMA AUTÓNOMO DE EVALUACIÓN DE COMPETENCIAS 50
B1. DESCRIPCIÓN DEL CASO DE USO 50
B2. ARQUITECTURA DEL SISTEMA DE IA 51
B3. INVENTARIO DE ACTIVOS 52
B4. TIPO DE SISTEMA DE IA 52
B5. RIESGOS ESPECÍFICOS IDENTIFICADOS 53
B6. APLICACIÓN DE MEDIDAS DE SEGURIDAD 55

1. INTRODUCCIÓN

1.1. Objeto

El Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad (en adelante ENS) [Ref.5], establece **los principios básicos** y **los requisitos mínimos** necesarios para una protección adecuada de la información tratada y de los servicios prestados por las entidades del sector público.

Cabe recordar que el ENS no solo se aplica a los **sistemas de información del sector público, sino también a aquellos del sector privado que, en virtud de una relación contractual, presten servicios competenciales a las entidades del sector público**, tal y como establece el artículo 2.3 del citado Real Decreto 311/2022. Asimismo, cualquier entidad, independientemente de su naturaleza jurídica, puede optar por aplicar voluntariamente el ENS como marco de referencia para mejorar la seguridad de sus sistemas.

La aparición y avance de la inteligencia artificial (IA) y su incorporación a los sistemas de información tiene un impacto significativo en la operación de las entidades tanto públicas como privadas. Los sistemas de IA pueden presentar un comportamiento probabilístico, generar resultados nuevos y, en algunos casos, actuar de forma autónoma, lo que puede dificultar la transparencia de las operaciones realizadas.

El objetivo de la presente guía es proporcionar un marco de referencia técnico y organizativo para el uso de sistemas de IA en los sistemas de información y servicios en el ámbito del ENS, incluidos los prestados por empresas del sector privado a entidades del sector público.

La guía aborda aspectos clave a lo largo de todo el ciclo de vida del sistema de IA en los servicios prestados por la entidad, haciendo especial énfasis en la gestión de riesgos, el marco de gobierno y la aplicación de las medidas de seguridad del ENS. Todo ello, en coherencia con las disposiciones del ENS, del Reglamento de inteligencia artificial (en adelante, RIA) [Ref.1] y del resto del marco normativo aplicable.

La presente guía se elabora en ejercicio de las competencias atribuidas al Centro Criptológico Nacional en materia de seguridad de los sistemas de información y de desarrollo del Esquema Nacional de Seguridad. Las referencias al RIA [Ref.1] se incorporan en la medida en que resultan pertinentes para el objeto y alcance de esta guía, sin que pretendan abordar de forma exhaustiva las obligaciones derivadas de dicho Reglamento ni establecer criterios generales para su interpretación o aplicación. En consecuencia, la observancia de las orientaciones y recomendaciones contenidas en esta guía no acredita, por sí sola, la conformidad con el RIA ni exime del cumplimiento de las obligaciones que de este se deriven. Todo ello se entiende sin perjuicio de las competencias atribuidas a la Agencia Española de Supervisión de Inteligencia Artificial (AESIA) y a las demás autoridades competentes en materia de aplicación y supervisión del RIA.

1.2. Alcance

La guía está dirigida a las entidades que desarrollen, utilicen, contraten, desplieguen o administren sistemas de IA en el ámbito de los sistemas de información y servicios sujetos al ENS. En este ámbito, resulta de especial interés para los responsables designados conforme al ENS que participen en la adquisición, diseño, configuración, operación o supervisión de los sistemas de IA.

1.3. Relación con el ENS y con otras guías del CCN

La presente guía se integra en el marco del ENS [Ref.5] y orienta a las entidades sobre la incorporación de los sistemas de IA a los procesos, requisitos y medidas de seguridad que este establece, sin crear un marco paralelo de gestión de la seguridad.

Asimismo, complementa la guía de buenas prácticas CCN-CERT BP/36 frente al modelo de IA ofensiva [Ref.8] y la guía CCN-STIC 861 Criterios de Aplicabilidad de las medidas del ENS frente a la IA ofensiva [Ref.7]. La guía CCN-CERT BP/36 analiza el impacto de la IA ofensiva en el escenario de amenazas y propone buenas prácticas para hacerle frente. La guía CCN-STIC 861, tomando como referencia la BP/36, concreta criterios verificables de aplicabilidad de medidas del anexo II del ENS frente a dicha amenaza. Esta guía, por su parte, se centra en el uso de sistemas de IA por la entidad y proporciona orientaciones específicas para quienes desarrollen, utilicen, contraten, desplieguen o administren sistemas de IA en el ámbito de los sistemas de información y servicios sujetos al ENS, sin modificar el catálogo de medidas ni su régimen de exigibilidad.

2. CONCEPTOS Y MODELO DE REFERENCIA

Según el RIA [Ref.1] y la guía 02 de definiciones de AESIA [Ref.10], se define un **sistema de inteligencia artificial** (en adelante, sistema de IA) como un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere a partir de la información de entrada los resultados de salida, como predicciones, contenidos, recomendaciones o decisiones.

Esta definición es lo más general posible para así **abarcar** el rango máximo de **tecnologías y modelos existentes y futuros**.

De acuerdo con el marco identificado, **un sistema de IA se caracteriza por su capacidad de inferir**, a partir de información de entrada, la forma de generar resultados de salida que pueden influir en entornos físicos o virtuales. Esta capacidad de inferencia, junto con sus distintos niveles de autonomía y posible capacidad de adaptación tras el despliegue, constituye uno de los principales elementos de diferenciación respecto del software tradicional basado exclusivamente en reglas definidas por personas físicas.

Teniendo en cuenta la definición del ENS de sistema de información como un grupo de dispositivos interconectados o relacionados entre sí, (por ejemplo, hardware, software, comunicaciones, instalaciones, personal) en el que uno o varios de ellos realicen, mediante un programa, el tratamiento automático de datos digitales. Un sistema de IA estaría compuesto por un **conjunto de estos activos**, pudiendo ser el sistema de IA una parte del sistema de información o el propio sistema de información, sobre el que se determina la conformidad con el ENS.

Una parte clave del sistema de IA sería el **modelo de IA**, que constituye el componente algorítmico entrenado, basado en datos o reglas lógicas. Su función principal es procesar entradas y generar resultados, patrones, predicciones, recomendaciones, contenidos o decisiones potenciales. El modelo de IA **no constituye un sistema de IA completo**, ya que necesita activos soporte (hardware, comunicaciones, etc.) para implementarse y ejecutarse.

A continuación, se detallan diferentes sistemas de IA según cómo operen o la tarea y el objetivo que persigan. Según su forma de operar, existen sistemas de IA, como por ejemplo:

- Un **agente** es un ejemplo de sistema de IA diseñado para funcionar con un comportamiento orientado a objetivos. Utiliza modelos de IA de uso general (como los LLM) como núcleo de razonamiento y se conecta con entornos o herramientas externas para percibir información, tomar decisiones adaptativas y ejecutar acciones de forma automatizada, operando bajo distintos niveles de autonomía sin requerir supervisión humana constante para cada tarea individual.
- Un **sistema de IA autónomo** es un ejemplo de sistema de IA que se refiere a cualquier modelo o software de IA capaz de ejecutar tareas, tomar decisiones o adaptarse a nuevos datos con una intervención humana mínima o nula (por

ejemplo, un algoritmo de recomendación que ajusta su contenido automáticamente, un filtro de spam con actuación o un modelo predictivo de precios con operación automática). Se debe tener en cuenta que no todos los sistemas de IA autónomos operan bajo una estructura de agentes con objetivos dinámicos, planificación y uso activo de herramientas.

Dependiendo de la tarea y el objetivo que persigan, existen varios ejemplos de sistema de IA como:

- Sistema de IA predictiva (riesgos, valores…).
- Asistente conversacional.

Debido a la gran variedad de componentes y relaciones entre ellos que puede tener un sistema de IA, resulta complejo establecer una única arquitectura tipo.

A modo de ejemplo, se muestra una arquitectura común de sistema de IA con la finalidad de identificar sus distintos componentes y sus relaciones. En este caso se trata de una arquitectura para un sistema de IA de peticiones con búsqueda de información adicional y capacidad de toma de decisiones. Cada componente (LLM, motor de búsqueda, aplicación externa, etc.) puede tener su propio hardware asociado y puede residir en los sistemas de la entidad o en los de un tercero. Los procesos transversales de seguridad deben abarcar todos los componentes independientemente de su ubicación.



Ilustración 1 Ejemplo de una arquitectura de un sistema de IA

Los componentes representados en la Ilustración 1 son los siguientes:

- **Usuario:** Origen de la consulta, petición o instrucción al sistema de IA, como puede ser una persona o una aplicación.
- **Filtros de entrada:** Mecanismos que validan, normalizan y analizan la entrada recibida, aplicando políticas de seguridad, clasificación o restricciones de uso.

- **Motor de búsqueda:** Componente que localiza y recupera la información más relevante desde las fuentes de conocimiento para dar contexto a la consulta.
- **Fuentes de conocimiento:** Repositorios de información utilizados por el sistema, como documentos, bases de datos, aplicaciones corporativas o fuentes externas autorizadas.
- **LLM:** Modelo de IA escogido por la entidad encargado de procesar la consulta y el contexto recuperado para generar una respuesta coherente y útil. Este modelo puede cambiar en función del contexto o de las preferencias del usuario.
- **Filtros de salida:** Controles aplicados sobre la respuesta generada para verificar aspectos de seguridad, privacidad, calidad, formato y cumplimiento de políticas.
- **Respuesta:** Resultado final entregado al usuario, que puede incluir texto, referencias, recomendaciones, advertencias, explicaciones, acciones o decisiones.
- **Otras aplicaciones o sistemas:** Aplicaciones, sistemas o servicios con los que el modelo interactúa para consultar información, ejecutar acciones o integrar funcionalidades adicionales.
- **Procesos transversales:** Capacidades de soporte que operan durante todo el ciclo de funcionamiento, como el registro de eventos, monitorización, control de accesos, gestión de incidentes, versionado y supervisión humana.

Por lo tanto, cada componente (modelo, datos, funcionalidades, etc.) de un sistema de IA puede tener un modelo de servicio diferente:

- Implantado en la entidad (implantado en los servidores de la entidad, sea desarrollado por la entidad o desarrollo externo).
- Proporcionado por un tercero (API, SaaS…).

Como ejemplo de implantación, una entidad puede contar con un sistema de IA que sea:

- Un agente con interfaces, bases de datos y varios modelos.
- Un modelo desarrollado en la entidad y un modelo desarrollado por un tercero.
- Una interfaz a una base de datos de la entidad, una interfaz a una base de datos externa y una interfaz a la aplicación de procesos de la entidad.
- Permisos sobre la aplicación de procesos para autorizar y gestionar procesos.

3. INVENTARIO DE ACTIVOS

La aparición de la IA conlleva a su vez la aparición de activos que conforman un sistema de IA. De igual manera, este apartado no se plantea como un inventario de activos nuevo e independiente, sino un ejercicio de definición de clases de activos que pueden ser incorporadas en el marco actual del ENS [Ref.5]. La entidad deberá identificar los activos que componen el sistema de IA, sus relaciones y su entorno.

Adicionalmente, es imprescindible entender que la IA es una tecnología en desarrollo constante y esta guía no puede incluir los avances futuros de la IA. Por lo que, sin perjuicio de que surjan nuevos activos en el futuro, se presentan las siguientes clases de activos, atendiendo siempre a la arquitectura del sistema de IA:

- **Activos de datos:** Engloban los diferentes tipos de conjuntos de datos utilizados en las distintas fases del sistema de IA como, por ejemplo:
  - Datos de entrenamiento
  - Datos de validación y prueba
  - Datos de producción o entrada
- **Activos de software, aplicación y modelos:** Contienen el modelo y sus elementos relacionados, entre ellos:
  - Modelos de IA entrenados
  - Librerías y frameworks de aprendizaje automático
  - Prompts de sistema
  - Parámetros de configuración
- **Activos de infraestructura y ejecución**: Los artefactos y recursos destinados al entrenamiento y evaluación como pueden ser:
  - Recursos de computación especializada (GPUs, NPUs…)
  - Entornos de entrenamiento, inferencia y ejecución
- **Activos de supervisión y trazabilidad:** Incluyen archivos de registros y elementos orientados a la supervisión de su comportamiento y a la trazabilidad. Por ejemplo:
  - Archivos de registro
  - Métricas de rendimiento del modelo
  - Interfaz humano-máquina
- **Activos externos en la cadena de valor**: Engloban los activos que provienen de terceros como pueden ser:
  - Modelos de IA de terceros
  - Conjunto de datos externos o licenciados
  - Sistemas de IA prestados por terceros

4. CLASIFICACIÓN Y GESTIÓN DE RIESGOS EN SISTEMAS DE IA

La gestión de los riesgos asociados a los sistemas de IA requiere abordar dos perspectivas complementarias. Por una parte, debe determinarse la clasificación regulatoria del sistema, atendiendo a su finalidad prevista, contexto de utilización y al impacto que puede producir. Por otra, resulta necesario identificar y evaluar los riesgos específicos derivados de su uso en la entidad. Ambas perspectivas permiten establecer un tratamiento proporcional, definir las obligaciones aplicables y determinar las medidas necesarias durante la utilización del sistema.

Es necesario actualizar el análisis de riesgos del sistema incorporando las amenazas, vulnerabilidades, dependencias y consecuencias asociadas al caso de uso del sistema de IA.

Los activos y riesgos específicos de los sistemas de IA se integrarán en la metodología ordinaria y no se gestionarán como un análisis aislado sin relación con los activos y servicios existentes.

El tratamiento podrá incluir medidas enfocadas a mitigar los riesgos propios de los sistemas de IA como limitaciones de alcance, controles técnicos y organizativos, supervisión humana, transparencia, aislamiento, contención, detención del sistema, reducción de permisos en tiempo real, alternativas operativas y condiciones contractuales. Cada medida deberá vincularse con el riesgo que pretende reducir y con el responsable de su implantación y seguimiento.

Los riesgos residuales, las incertidumbres relevantes y el plan de tratamiento deberán ser asumidos formalmente por la entidad siguiendo su política de seguridad. La evaluación se revisará ante cambios significativos, incidentes, cambio de proveedor, nuevos conocimientos sobre amenazas o desviaciones respecto del comportamiento previsto.

4.1. Clasificación de sistemas de IA conforme al RIA

Basándose en el marco normativo expuesto anteriormente, el RIA [Ref.1] introduce una clasificación de riesgos en función de la finalidad y ámbito del sistema de IA. Las entidades deberán tener en cuenta este apartado para estudiar si su sistema de IA se encuentra representado o no en esta clasificación.

- **Prácticas de IA prohibidas (art.5)**

  El RIA prohíbe taxativamente la introducción en el mercado, puesta en servicio o utilización de sistemas de IA cuyas prácticas se consideran contrarias a los valores de la Unión Europea y a los derechos fundamentales. Estas prácticas prohibidas son, entre otras:

  - Técnicas manipuladoras o subliminales. Sistemas que trasciendan la conciencia de una persona o utilicen técnicas engañosas para alterar sustancialmente su comportamiento, mermando su capacidad de tomar una decisión informada provocando o pudiendo provocar perjuicios considerables.

- Explotación de vulnerabilidades humanas. Sistemas que se aprovechen de la edad, discapacidad o situación social o económica específica de una persona o colectivo para alterar su comportamiento causándole perjuicios considerables.
- Predicción de delitos. Realizar evaluaciones de riesgo para predecir la probabilidad de que una persona física cometa un delito basándose únicamente en la elaboración de su perfil o en sus rasgos de personalidad.
- Bases de datos faciales masivas. Crear o ampliar bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión.
- Inferencia de emociones en el trabajo o educación. Reconocer emociones en lugares de trabajo o centros educativos, excepto cuando se instalen por motivos médicos o de seguridad.
- Categorización biométrica sensible. Clasificar a personas basándose en datos biométricos para deducir su raza, opiniones políticas, afiliación sindical, convicciones religiosas o filosóficas, vida u orientación sexuales.
- Identificación biométrica remota "en tiempo real" en espacios públicos. Su uso con fines policiales está prohibido, de modo que ninguna autoridad podrá adoptar decisiones adversas basándose exclusivamente en sus resultados, salvo en excepciones muy concretas limitadas y con autorización previa.

- **Alto riesgo (art.6)**

  El RIA define las reglas de clasificación para sistemas de alto riesgo:

  - Sistemas de IA que forman parte de un componente de seguridad en productos del anexo I del RIA
  - Sistemas de IA que operan en alguno de los ámbitos del anexo III del RIA.

- **Obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA (art.50)**

  Con independencia de las clasificaciones anteriores, determinados sistemas pueden quedar sujetos a obligaciones específicas de transparencia para que las personas puedan tomar decisiones informadas:

  - Interacción con humanos. Las personas deben ser informadas de que están interactuando con un sistema de IA (como chatbots, salvo que sea evidente por contexto).
  - Generación y manipulación de contenidos (Deepfakes). Los sistemas (incluidos los de IA generativa) que generen contenidos de imagen, audio, vídeo o texto de forma sintética deben revelar que han sido generados o manipulados artificialmente, y marcar dichos resultados en un formato legible por máquina.

- Reconocimiento de emociones y categorización biométrica. Los responsables de su uso deben informar de su funcionamiento a las personas físicas expuestas a él.

Estas obligaciones resultan aplicables tanto a sistemas de IA de alto riesgo como a sistemas que no entren en dicha clasificación.

- **Reglas de clasificación de los modelos de IA de uso general como modelos de IA de uso general con riesgo sistémico (art.51)**

  El RIA introduce una categoría especial para los "modelos de IA de uso general" (que presentan gran generalidad y son capaces de realizar de manera competente una gran variedad de tareas, como los grandes modelos de lenguaje) que constituyen un régimen específico del Capítulo V.

  Dentro de estos, se clasifican con riesgo sistémico (riesgos graves de perturbación de sectores críticos, procesos democráticos o difusión de contenidos ilícitos a gran escala) si:

  - La Comisión decide que tienen capacidades de gran impacto evaluadas con metodologías y parámetros de referencia técnicos, basándose en los criterios establecidos en el anexo XIII del RIA.
  - Si la cantidad acumulada de cálculo utilizada para su entrenamiento es superior a un número determinado de operaciones de coma flotante (FLOP).

  Según el considerando 100 del RIA, cuando un modelo de IA de uso general esté integrado en un sistema de IA, este sistema de IA deberá considerarse sistema de IA de uso general.

4.2. Riesgos asociados al uso de sistemas de IA

El uso de sistemas de IA introduce nuevos riesgos en la entidad. La comprensión de dichos riesgos resulta fundamental para establecer medidas proporcionales al contexto de utilización y al impacto potencial del sistema.

La identificación de los riesgos recogidos a continuación se ha realizado tomando como referencia marcos especializados en seguridad de sistemas de IA, destacando OWASP Top 10 for LLM Applications 2026 [Ref.9] o la guía 11 de ciberseguridad AESIA [Ref.11], así como las disposiciones establecidas en el RIA en materia de gestión de riesgos, supervisión humana, gobernanza, trazabilidad y protección de los derechos fundamentales. Debe tenerse en cuenta que marcos de referencia como OWASP son objeto de revisiones y actualizaciones periódicas para adaptarse a la evolución tecnológica y a la aparición de nuevos riesgos.

- **Divulgación y tratamiento inadecuado de información (R01)**
  - **Dimensión afectada: Confidencialidad**

  Se refiere al riesgo de que la información proporcionada o accesible al sistema de IA sea revelada, conservada, reutilizada o tratada en condiciones no

autorizadas. Puede afectar tanto a los datos introducidos mediante prompts y archivos adjuntos como a las respuestas, historiales, memorias y demás información generada durante el uso del sistema. Ejemplos:

- Introducción de datos personales, información sensible o documentación interna en un servicio no autorizado.
- Conservación o utilización de prompts, archivos y respuestas para mejorar o entrenar modelos sin conocimiento o autorización de la entidad.

- **Acceso indebido a la información (R02)**
  - **Dimensión afectada: Confidencialidad, integridad y autenticidad.**

  Se produce cuando el sistema de IA accede o facilita el acceso a información que no resulta necesaria para el caso de uso autorizado, debido a permisos excesivos, repositorios mal configurados, conectores inadecuados o una aplicación insuficiente del principio de mínimo privilegio. Ejemplos:

  - Recuperación y síntesis de documentos almacenados en repositorios corporativos cuyos permisos no han sido revisados previamente.
  - Configuración de un conector con privilegios que permiten consultar y alterar información perteneciente a unidades o a grupos distintos de los autorizados.

- **Resultados o decisiones incorrectas o no fiables (R03)**
  - **Dimensión afectada: Integridad y disponibilidad.**

  Comprende la generación de información falsa, incompleta, desactualizada, inconsistente o insuficientemente fundamentada presentada con un elevado grado de convicción o con apariencia de veracidad (fenómeno conocido habitualmente como alucinaciones). Este riesgo puede provocar errores cuando los resultados se utilizan sin comprobar sus fuentes, limitaciones o contexto de aplicación. Ejemplos:

  - Generación de referencias normativas, resoluciones o fuentes documentales inexistentes.
  - Elaboración de un borrador jurídico que incorpora citas incorrectas o conclusiones que no se corresponden con la normativa aplicable.

- **Manipulación de instrucciones, contexto y fuentes de conocimiento (R04)**
  - **Dimensión afectada: Todas**

  Se refiere a la alteración del comportamiento previsto del sistema mediante instrucciones o contenidos diseñados para modificar sus salidas, logrando eludir sus controles de seguridad (jailbreaking) o alterar sus instrucciones de sistema (system prompt). Esta manipulación puede producirse de forma directa por un usuario o indirectamente a través de documentos, páginas web, correos electrónicos o fuentes de conocimiento utilizadas por el sistema. Ejemplos:

- Inclusión de instrucciones ocultas en un documento para que el sistema ignore sus reglas y revele información del contexto o de sus instrucciones internas.
- Incorporación de contenido malicioso en una fuente documental para influir deliberadamente en las salidas proporcionadas por el sistema.

- **Uso indebido de conectores, herramientas y capacidades de actuación (R05)**
  - **Dimensión afectada: Todas**

Se produce cuando el sistema utiliza conectores, aplicaciones, APIs u otras herramientas para acceder a información o ejecutar acciones no autorizadas, innecesarias o desproporcionadas. El riesgo aumenta cuando el sistema dispone de permisos elevados, puede encadenar varias acciones o puede interactuar con otros sistemas de forma autónoma. Ejemplos:

- Envío de una comunicación, modificación de un registro o eliminación de información sin una validación humana previa.
- Utilización de credenciales con privilegios excesivos para consultar o actuar sobre sistemas corporativos conectados.

- **Resultados sesgados o no representativos (R06)**
  - **Dimensión afectada: Todas**

Comprende la generación, recomendación o utilización de resultados condicionados por sesgos en los datos, errores en el modelo, deficiencias en las fuentes de información o limitaciones del contexto proporcionado. Como consecuencia, el sistema puede producir respuestas no objetivas, inconsistentes o alejadas de la realidad, afectando a la fiabilidad de los procesos y decisiones que dependen de él. Este riesgo adquiere especial relevancia cuando los sistemas de IA se emplean para apoyar tareas de análisis, evaluación, supervisión o toma de decisiones dentro de los sistemas de información. Ejemplos:

- Emisión de conclusiones inexactas, parciales o no suficientemente fundamentadas que condicionen decisiones operativas o de gestión.
- Utilización de una respuesta sesgada como fundamento para valorar solicitudes, priorizar expedientes o adoptar decisiones.

- **Falta de supervisión, trazabilidad y control organizativo (R07)**
  - **Dimensión afectada: Todas**

Se refiere a los riesgos derivados de una utilización insuficientemente supervisada de los sistemas de IA, de ausencia de responsables claramente identificados o de la imposibilidad de conocer cómo se ha obtenido un resultado o realizado una actuación. Ejemplos:

- Ausencia de registros suficientes para auditar una actuación o investigar un incidente.

- Utilización de sistemas de IA fuera de los procedimientos autorizados por la entidad.

- **Dependencia de terceros y cadena de suministro (R08)**

  - **Dimensión afectada: Disponibilidad y trazabilidad**

  Comprende aquellos riesgos derivados de la utilización de sistemas de IA y sus componentes proporcionados por terceros, sobre los que la entidad no dispone de un control completo. Esta dependencia puede afectar a la continuidad del servicio, la transparencia del funcionamiento del sistema o la capacidad de verificar el cumplimiento de las medidas de seguridad. Ejemplos:

  - Modificación del comportamiento de un modelo como consecuencia de una actualización realizada por el proveedor.
  - Interrupción o degradación del servicio debido a incidencias en infraestructuras externas.

5. MARCO DE GOBIERNO

Es necesario formar un marco robusto para hacer frente a los riesgos derivados de la IA expuestos anteriormente. Para ello, se parte de la base ya establecida por el ENS y sus guías, incluyendo así los principios adicionales expuestos en el RIA. Los sistemas de IA introducen nuevas responsabilidades, riesgos y obligaciones que deben integrarse en los diferentes bloques de gobierno y dirección, supervisión y operación, facilitando así la aplicación práctica del marco normativo especificado.

5.1. Principios generales

El ENS [Ref.5] define una serie de principios básicos en el artículo 5 (seguridad como proceso integral, vigilancia continua, etc.). Estos principios son la base para la gestión de la seguridad, pero el nuevo entorno de la IA requiere complementarlos con los siguientes principios:

- **Principio 1. Supervisión humana**

  Los sistemas de IA se diseñarán y desarrollarán de modo que puedan ser vigilados de manera efectiva por personas físicas durante el período que estén en uso, lo que incluye dotarlos de herramientas de interfaz humano-máquina adecuadas. Existen dos técnicas reconocidas para garantizar este principio:

  - Human-in-the-loop (HITL): El sistema no puede avanzar ni ejecutar una acción sin la aprobación o intervención directa de la persona en cada paso crítico.
  - Human-on-the-loop (HOTL): El sistema actúa y ejecuta por sí solo, mientras la persona supervisa desde fuera como una red de seguridad. El humano no interviene en cada paso, pero vigila el proceso general y puede detenerlo, corregirlo o intervenir solo si detecta una anomalía o un error.

- **Principio 2. Transparencia y trazabilidad.**

  Las entidades deben ser capaces de identificar los sistemas utilizados, sus finalidades, las fuentes de información empleadas y las actuaciones realizadas durante su operación. La existencia de mecanismos de trazabilidad facilita la supervisión, la auditoría, la investigación de incidentes y la rendición de cuentas, además de contribuir a generar confianza en la utilización de estas tecnologías.

- **Principio 3. Mitigación de sesgos**

  Es imperativo que el responsable de los datos de entrenamiento audite activamente los datos de entrenamiento y los resultados para evitar sesgos discriminatorios por razones de género, raza, edad o nivel socioeconómico, garantizando la igualdad de trato.

5.2. Integración en el marco de gobierno

Las entidades podrán actuar, **según la posición que ocupen en la cadena de valor** y la actividad que desarrollen, como proveedores, responsables del despliegue, importadores, distribuidores, representantes autorizados, fabricantes de productos o proveedores de modelos de IA de uso general. Una misma entidad podrá ocupar distintas posiciones respecto de diferentes sistemas de IA o, incluso, acumular varias posiciones respecto de un mismo sistema cuando concurran las circunstancias previstas en el RIA [Ref.1].

Estas figuras deben entenderse como **posiciones jurídicas a nivel de entidad**, de las que derivan obligaciones específicas, y no como cargos internos equivalentes a los previstos en el ENS [Ref.5] y en la guía CCN-STIC 801 [Ref.6]. Por tanto, resulta necesario que cada entidad determine, dentro de su modelo de gobernanza, las **responsabilidades internas** necesarias para poder asumir con garantías las obligaciones correspondientes a cada posición jurídica.

En consecuencia, una vez identificado el contexto y los sistemas de IA afectados, corresponde a cada entidad:

- Determinar la posición o posiciones jurídicas que ocupa respecto de cada sistema o modelo de IA.
- Clasificar los sistemas conforme a la clasificación de riesgos definida en el RIA [Ref.1] y en el apartado 4.1.
- Identificar las obligaciones aplicables.
- Identificar las nuevas obligaciones de seguridad de información derivadas del uso de IA.
- Traducir dichas obligaciones a su modelo organizativo.
- Distribuirlas entre los bloques de gobierno y dirección, supervisión y operación.
- Establecer mecanismos de coordinación con las funciones de seguridad, protección de datos, cumplimiento, contratación, asesoría jurídica, control interno y gestión de riesgos.

A continuación, se propone una organización de dichas obligaciones en función de los distintos bloques de responsabilidad en la entidad:

5.2.1. Gobierno y dirección

Corresponde al bloque de gobierno y dirección la adopción de las decisiones estratégicas relativas al desarrollo, adquisición, puesta en servicio y utilización de sistemas de IA, así como la aprobación del marco general de gobernanza necesario para garantizar el cumplimiento del RIA [Ref.1].

En este bloque deberían situarse, entre otras, las siguientes obligaciones:

- Aprobar la política corporativa de gobernanza y uso responsable de la IA.

- Modificar la política de seguridad de la información para incluir los nuevos principios, los nuevos riesgos que introducen los sistemas de IA, las funciones de seguridad derivadas del uso de sistemas de IA y su asignación a los roles o comités correspondientes.
- Aprobar la normativa de la entidad para tener en cuenta el uso de los sistemas de IA. En particular, la norma de uso de los sistemas de IA por parte de los empleados, que podrá integrarse en la norma de uso de medios electrónicos.
- Determinar el nivel de tolerancia al riesgo asociado al uso de sistemas de IA.
- Aprobar los criterios para identificar y clasificar los sistemas y modelos de IA.
- Determinar la posición jurídica que corresponde a la entidad respecto a su sistema de IA.
- Aprobar la introducción o utilización de sistemas de IA de alto riesgo y de sistemas de IA en función de la información del inventario del sistema de IA y, en su caso, del piloto de pruebas realizado.
- Asignar los recursos humanos, técnicos, económicos y organizativos necesarios.
- Asegurar, cuando resulte exigible, la realización y validación interna de las evaluaciones de impacto relativas a los derechos fundamentales cuya responsabilidad corresponde a los responsables del despliegue comprendidos en el artículo 27 del RIA [Ref.1].
- Asegurar la implantación de mecanismos de supervisión humana y que las personas designadas dispongan de autoridad, competencia, formación y recursos adecuados.
- Aprobar las condiciones bajo las cuales puedan realizarse modificaciones, adaptaciones o cambios de finalidad en los sistemas de IA.
- Establecer criterios para la suspensión, retirada o sustitución de sistemas que no sean conformes o presenten un riesgo inaceptable.

Cuando la entidad actúe como proveedor de sistemas de IA de alto riesgo, el bloque de gobierno y dirección deberá garantizar los refuerzos establecidos por la normativa vigente.

5.2.2. Supervisión

Corresponde a este bloque la función de asesoramiento, propuesta, coordinación, control y supervisión del cumplimiento del RIA [Ref.1], de la seguridad de los sistemas de IA y de la adecuación de las medidas implantadas.

Se propone que el Responsable de la Seguridad participe en este bloque respecto de las funciones vinculadas con la seguridad de los sistemas de información y, en particular, en las siguientes:

- Identificar los sistemas de información que soportan o integran sistemas de IA.

- Elaborar la normativa de la entidad para tener en cuenta el uso de los sistemas de IA.
- Asesorar en la introducción y autorización de sistemas de IA de alto riesgo y de sistemas de IA en función de la información del inventario del sistema de IA y, en su caso, del piloto de pruebas realizado.
- Determinar el ámbito de aplicación del ENS [Ref.5] y de la categoría correspondiente.
- Trasladar los riesgos derivados de la IA al análisis de riesgos de seguridad.
- Proponer requisitos de seguridad aplicables a la información, los servicios, los modelos, los datos y las infraestructuras.
- Supervisar la implantación de medidas de exactitud, solidez, resiliencia y ciberseguridad.
- Coordinar la gestión de vulnerabilidades, incidentes y continuidad.
- Supervisar los mecanismos de registro, trazabilidad y conservación de evidencias.
- Valorar los riesgos asociados a proveedores, servicios en la nube, modelos externos, interfaces de programación y componentes de terceros.
- Evaluar el impacto de las modificaciones del sistema sobre su seguridad.
- Coordinarse con las funciones responsables de protección de datos, cumplimiento y gestión de riesgos.

Y también, obligaciones derivadas del RIA como pueden ser:

- Asesorar sobre la clasificación de los sistemas de IA y, en particular, sobre su consideración como prácticas prohibidas, de alto riesgo o sujetos a obligaciones específicas de transparencia.
- Supervisar el inventario de sistemas y modelos de IA.
- Coordinar o prestar apoyo a la elaboración de la evaluación de impacto relativa a los derechos fundamentales cuando resulte exigible, sin perjuicio de que la obligación jurídica establecida en el artículo 27 del RIA [Ref.1] corresponda al responsable del despliegue en los supuestos previstos en dicho artículo.
- Verificar que las personas designadas para la supervisión humana prevista en el artículo 14 del RIA, que es distinta de la supervisión organizativa de la seguridad establecida en el ENS, disponen de competencia, formación, autoridad e independencia suficientes.
- Verificar el cumplimiento de los requisitos de información y transparencia.
- Coordinar la gestión de reclamaciones, incidentes graves y situaciones de no conformidad.

- Asesorar a los órganos de contratación sobre los requisitos que deben incorporarse en pliegos, contratos y acuerdos.
- Supervisar los cambios realizados sobre los sistemas de IA y valorar cuándo pueden constituir una modificación sustancial.
- Coordinar la cooperación con las autoridades competentes y los organismos de evaluación de la conformidad.

5.2.3. Operación

Corresponde a este bloque la ejecución material de las actividades de diseño, desarrollo, adquisición, puesta en servicio, utilización, mantenimiento y retirada de los sistemas de IA derivadas de las decisiones internas, normas y procedimientos de seguridad y de uso de sistemas de IA elaborados y aprobados por la entidad.

Cuando la entidad actúe como **proveedor de un sistema** de IA de alto riesgo, este bloque deberá ejecutar las medidas necesarias, entre otras:

- Aplicar el sistema de gestión de riesgos durante todo el ciclo de vida.
- Garantizar la gobernanza y calidad de los datos.

Cuando la entidad actúe como **responsable del despliegue** de un sistema de IA de alto riesgo, este bloque deberá asegurar, entre otros aspectos, que:

- El sistema se utiliza conforme a las instrucciones del proveedor.
- Los datos de entrada bajo el control de la entidad son pertinentes y suficientemente representativos para la finalidad prevista.

5.3. Formación, concienciación y alfabetización en IA

La formación y la concienciación deben alinearse con el artículo 4 del RIA [Ref.1] que introduce una **obligación horizontal de alfabetización en IA**, adoptando medidas para apoyar la promoción de la alfabetización en materia de IA en la entidad. Esto englobará la formación de los usuarios y demás personas que se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA, teniendo en cuenta sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso de los sistemas de IA y las personas o los colectivos de personas en que se van a utilizar dichos sistemas. Esta obligación no exige que los proveedores o los responsables del despliegue garanticen un nivel específico de alfabetización en materia de IA de ninguna persona en particular.

Ello complementará las obligaciones de formación y concienciación definidas en las medidas de formación y concienciación del anexo II del ENS [Ref.5].

6. IMPLANTACIÓN DE LA SEGURIDAD

El uso de sistemas de IA supone un cambio profundo en la gestión de la seguridad, especialmente en la configuración, mantenimiento, operación y supervisión de los sistemas.

En caso de que el uso de un sistema de IA afecte al inventario de información y servicios, será necesario actualizar la valoración de la información y servicios y la declaración de aplicabilidad. A su vez, es necesario tener en cuenta los nuevos principios generales especificados en el apartado 5.1.

6.1. Criterios de aplicabilidad de las medidas del ENS

A continuación, se presenta una serie de criterios de aplicabilidad de las medidas del anexo II del ENS [Ref.5] incorporando consideraciones específicas para sistemas de IA. Estos criterios tienen carácter orientativo y flexible, debiendo adaptarse a la finalidad prevista del sistema, a su contexto de utilización, a su arquitectura y a los riesgos específicos identificados por la entidad en el análisis de riesgos.

Los criterios establecidos en esta guía no determinan por sí mismos la exigibilidad de una medida del ENS. Esta vendrá determinada por la categoría del sistema, las dimensiones de seguridad, los niveles correspondientes y, cuando proceda, el análisis de riesgos.

La selección no es exhaustiva: las medidas no incluidas seguirán siendo aplicables cuando corresponda conforme al ENS. El uso de un sistema de IA no eleva automáticamente la categoría del sistema.

6.1.1. [org.1] Política de seguridad

Código	Medida del ENS
org.1	Política de seguridad
En la política de seguridad de la información (PSI) se integrarán los nuevos principios, los nuevos riesgos, las funciones de seguridad y su asignación a los roles, según lo especificado en el apartado 5.2.	

6.1.2. [org.2] Normativa de seguridad

Código	Medida del ENS
org.2	Normativa de seguridad
Deben establecerse normas específicas respecto a la implantación y el uso de sistemas de IA en la entidad, definiendo los usos permitidos y prohibidos, regulando el uso de las herramientas de IA corporativa en todo su ciclo de vida (entrada, interacciones, salida, etc.).	

6.1.3. [org.3] Procedimientos de seguridad

Código	Medida del ENS
org.3	Procedimientos de seguridad
Se actualizarán los procedimientos de seguridad para detallar de forma clara y precisa cómo asegurar el uso de sistemas de IA de la entidad. En caso necesario, se elaborarán procedimientos adicionales.	

6.1.4. [org.4] Proceso de autorización

Código	Medida del ENS
org.4	Proceso de autorización
El sistema de IA y sus componentes deberán integrarse en el proceso formal de autorizaciones de la entidad. Además, se establecerán mecanismos para la aprobación de los nuevos sistemas de IA por el bloque de gobierno y dirección de la entidad, incluyendo la revisión de toda la información proporcionada del sistema de IA (finalidad, usuario, interfaz, datos…) y evaluando el nivel de riesgo.	

6.1.5. [op.pl.1] Análisis de riesgos

Código	Medida del ENS
op.pl.1	Análisis de riesgos
En el proceso de análisis y gestión de riesgos se tendrán en cuenta adicionalmente aquellos activos, amenazas y riesgos residuales derivados del uso de los sistemas de IA. Se han definido clases de activos y riesgos en los apartados 3 y 4.	

6.1.6. [op.pl.2] Arquitectura de seguridad

Código	Medida del ENS
op.pl.2	Arquitectura de seguridad
En la arquitectura de seguridad se incluirán los sistemas de IA, documentando el sistema de IA, su funcionalidad, activos y relaciones.	

6.1.7. [op.pl.3] Adquisición de nuevos componentes

Código	Medida del ENS
op.pl.3	Adquisición de nuevos componentes

En el proceso de adquisición se considerarán los componentes del sistema de IA (acceso a modelos, bases de datos, agentes, infraestructura, etc.)

6.1.8. [op.pl.4] Dimensionamiento / Gestión de la capacidad

Código	Medida del ENS
op.pl.4	Dimensionamiento / Gestión de la capacidad
Se tendrán en cuenta los requisitos de computación, almacenamiento, personal y comunicaciones específicos de los sistemas de IA.	

6.1.9. [op.acc.1] Identificación

Código	Medida del ENS
op.acc.1	Identificación
Se tendrán en cuenta las identidades de los diferentes componentes de los sistemas de IA (modelos, agentes…), permitiendo su identificación y trazabilidad.	

6.1.10. [op.acc.2] Requisitos de acceso

Código	Medida del ENS
op.acc.2	Requisitos de acceso
Se asignarán los permisos estrictamente necesarios en los componentes (modelos, agentes, interfaces…) del sistema de IA para el desempeño de sus funciones.	

6.1.11. [op.acc.3] Segregación de funciones y tareas

Código	Medida del ENS
op.acc.3	Segregación de funciones y tareas
La segregación de funciones implementada en la entidad se trasladará a los sistemas de IA cuando estos realicen tareas críticas como desarrollo, operación, autorización, etc.	

6.1.12. [op.acc.4] Proceso de gestión de derechos de acceso

Código	Medida del ENS
op.acc.4	Proceso de gestión de derechos de acceso

Se controlará y monitorizará los accesos otorgados a los sistemas de IA y sus componentes, cumpliendo el principio de mínimo de privilegio y necesidad de conocer.

6.1.13. [op.exp.1] Inventario de activos

Código	Medida del ENS
op.exp.1	Inventario de activos
En el inventario de activos, estarán debidamente identificados todos los componentes del sistema de IA, teniendo en cuenta lo definido en el apartado 3.	

6.1.14. [op.exp.2] Configuración de seguridad

Código	Medida del ENS
op.exp.2	Configuración de seguridad
En la configuración de seguridad de los sistemas de IA, es necesario implementar medidas técnicas, operacionales y en tiempo de ejecución (guardrails o guardarraíles de IA) que validen, filtren y acoten las entradas, procesos y salidas del sistema de IA para mantener su comportamiento seguro, conforme y dentro del inventario o caso de uso autorizado. Estas medidas adicionales se clasifican en: • Entrada: Evalúan los datos de entrada antes de que lleguen al modelo del sistema de IA. Por ejemplo: bloqueo de ataques de inyección y detección o filtración de datos personales o información confidencial. • Salida: Inspeccionan y controlan las salidas generadas por el sistema de IA o sus elementos (modelos). Por ejemplo: verificación de alucinaciones o filtros de sesgos. • De acción: Regulan las herramientas, permisos e interfaces de API que el sistema de IA ejecuta en tiempo real. Por ejemplo: aislamiento del sistema evitando conexiones a internet, restricción de permisos o activación de supervisión humana.	

6.1.15. [op.exp.3] Gestión de la configuración de seguridad

Código	Medida del ENS
op.exp.3	Gestión de la configuración de seguridad
Se gestionarán las medidas técnicas, operacionales y en tiempo de ejecución definidas en la configuración de seguridad.	

6.1.16. [op.exp.5] Gestión de cambios

Código	Medida del ENS
op.exp.5	Gestión de cambios
En el proceso de gestión de cambios se integrará el sistema de IA haciendo especial énfasis en los cambios de los componentes del sistema de IA (como modelos) que puedan tener repercusiones en la operación y calidad del sistema.	

6.1.17. [op.exp.7] Gestión de incidentes

Código	Medida del ENS
op.exp.7	Gestión de incidentes
Se tendrá en cuenta: • En la fase de detección, la monitorización de los componentes del sistema de IA (por ejemplo, variaciones en los resultados esperados). • En la fase de análisis y evaluación, los requisitos de transparencia del sistema. • En la fase de contención, la necesidad de supervisión humana, detención del sistema, etc.	

6.1.18. [op.exp.8] Registro de la actividad

Código	Medida del ENS
op.exp.8	Registro de la actividad
La entidad debe tener registros específicos de las actividades de los componentes del sistema de IA y procesos que permitan que los responsables interpreten y usen correctamente sus resultados de salida. Las actividades de los usuarios o procesos del sistema de IA se integrarán en los registros de auditoría existentes o se generarán registros adicionales si fuera necesario.	

6.1.19. [op.nub.1] Protección de servicios en la nube

Código	Medida del ENS
op.nub.1	Protección de servicios en la nube
Se prestará especial atención en el caso de que el sistema de IA se preste desde un servicio en la nube, estableciendo medidas de seguridad en función del modelo de servicio que se utilice. Cuando se utilicen servicios en la nube suministrados por terceros estos deberán ser	

conformes con el ENS o cumplir las medidas desarrolladas por las guías CCN-STIC relativas a la protección de servicios de IA en la nube.

6.1.20. [op.ext.1] Contratación y acuerdos de nivel de servicio

Código	Medida del ENS
op.ext.1	Contratación y acuerdos de nivel de servicio
En la contratación de un componente del sistema de IA prestado o desarrollado por un tercero, se le trasladarán requisitos de transparencia y de colaboración en la gestión de la seguridad del componente (p. ej. información de configuración, información de incidentes, información de accesos, niveles de servicio, etc.).	

6.1.21. [op.ext.3] Protección de la cadena de suministro

Código	Medida del ENS
op.ext.3	Protección de la cadena de suministro
Se analizará el impacto que puede tener sobre el sistema un incidente accidental o deliberado que tenga su origen en un proveedor del sistema de IA y esta dependencia se tendrá en cuenta en el plan de continuidad de la entidad. Adicionalmente, se deberán cumplir las obligaciones impuestas en el RIA según la posición jurídica que la entidad desarrolle en la cadena de valor (responsable del despliegue, proveedor…) para garantizar la seguridad del sistema de IA.	

6.1.22. [op.ext.4] Interconexión de sistemas

Código	Medida del ENS
op.ext.4	Interconexión de sistemas
Todos los intercambios de información dentro del sistema de IA y sus componentes deberán ser autorizados previamente. Asimismo, se tendrán que autorizar los flujos de prestación de servicios con otros sistemas o aplicaciones a los que tenga acceso el sistema de IA.	

6.1.23. [op.mon.1] Detección de intrusión

Código	Medida del ENS
op.mon.1	Detección de intrusión
Se reforzarán las herramientas de detección y prevención de intrusiones sobre el sistema de IA, especialmente en la interacción con componentes externos. Las herramientas de	

detección contemplarán las entradas al sistema, las salidas y/o las acciones que realice el sistema de IA.

6.1.24. [op.mon.3] Vigilancia

Código	Medida del ENS
op.mon.3	Vigilancia
Es necesario monitorizar el cumplimiento de las medidas adicionales de configuración de seguridad, así como los usuarios activos, volúmenes, funciones utilizadas, fuentes consultadas, herramientas invocadas, acciones, errores, bloqueos, consumo, disponibilidad, conexiones y eventos anómalos. Los indicadores deberán permitir detectar desviaciones, crecimiento no previsto, usos residuales o dependencia creciente. Se definirá qué indicadores pueden activar la intervención humana. Cuando se activen los indicadores que pueden hacer necesaria la intervención humana (Human-on-the-loop), se desarrollará el siguiente proceso: • Evaluar los indicadores que han activado la intervención humana (datos de entrada sospechosos, posibles conexiones ilícitas…). • Determinar la decisión a adoptar: no intervenir, corregir o detener el sistema. • Ejecutar la decisión y documentarla. La frecuencia de la monitorización se ajustará a la criticidad y podrá intensificarse tras la puesta en servicio, cambios o incidentes. Los hallazgos relevantes se comunicarán a los responsables y se incorporarán a las revisiones. En procesos críticos se evaluará la posibilidad de introducir puntos de control obligatorios de supervisión humana en el proceso ("Human-in-the-loop").	

6.1.25. [mp.per.3] Concienciación

Código	Medida del ENS
mp.per.3	Concienciación
Para cumplir el principio definido de alfabetización de IA, se trabajará en concienciar regularmente al personal acerca del uso seguro del sistema de IA.	

6.1.26. [mp.per.4] Formación

Código	Medida del ENS
mp.per.4	Formación
Se deberá formar a los usuarios y al personal que trabajen con los sistemas de IA para garantizar un nivel suficiente de conocimiento y comprensión sobre la IA.	

6.1.27. [mp.sw.1] Desarrollo de aplicaciones

Código	Medida del ENS
mp.sw.1	Desarrollo de aplicaciones
En el caso de desarrollo de sistemas de IA, se seguirán las pautas de desarrollo vigentes en la entidad (separación de entornos, desarrollo seguro, datos de pruebas). Dentro del ciclo de vida del desarrollo de un sistema de IA se tendrán en cuenta la prevención de envenenamiento de datos de entrenamiento, la anonimización de datos, la validación de librerías, la realización de pruebas de ataque adversario, etc. En el caso del desarrollo de aplicaciones asistido por IA, será necesaria una supervisión humana y pruebas adicionales sobre el contenido generado por el sistema de IA.	

6.1.28. [mp.info.1] Datos personales

Código	Medida del ENS
mp.info.1	Datos personales
Es necesario asegurar que los sistemas de IA tengan en cuenta la normativa de protección de datos en todo su ciclo de vida (entrenamiento, desarrollo, implantación) y cuenten con refuerzos que prevengan la exfiltración de datos personales.	

6.1.29. [mp.info.6] Copia de seguridad

Código	Medida del ENS
mp.info.6	Copia de seguridad
El procedimiento de copias de seguridad de la entidad deberá integrar el sistema de IA y todos sus componentes para poder recuperarse de un incidente. Los plazos de retención serán definidos en la normativa interna de la entidad.	

6.1.30. [mp.s.2] Protección de servicios y aplicaciones web

Código	Medida del ENS
mp.s.2	Protección de servicios y aplicaciones web
Para aquellos servicios que incorporen funcionalidades basadas en sistemas de IA, se considerarán adicionalmente las amenazas específicas asociadas a este tipo de sistemas (prompt injection, generación de respuestas no autorizadas, evasión de las restricciones de seguridad establecidas para el sistema, etc.).	

6.2. Matriz de escenarios y medidas relacionadas

La siguiente matriz facilita una propuesta de trazabilidad entre los escenarios de aplicación y las medidas de seguridad. No sustituye al análisis de riesgos ni constituye una lista cerrada.

Cabe destacar que la aplicación y refuerzo de las medidas se determinarán según la finalidad prevista del sistema, su categoría, su contexto de utilización, su arquitectura y los riesgos específicos identificados en el análisis de riesgos.

Las medidas de la Tabla 1 se aplican de forma acumulativa:

- Las medidas del escenario **Transversal** son aplicables a cualquier sistema de IA, según su finalidad prevista, su categoría, su contexto de utilización y su arquitectura.
- Las medidas del escenario **Alto riesgo y/o sistema autónomo** son aplicables, adicionalmente a las transversales, por la clasificación de sistemas de IA del apartado 4.1 y/o definición de sistema autónomo en el apartado 2.
- Las medidas del escenario **Riesgos específicos identificados** son aplicables, adicionalmente a las transversales, de forma selectiva, según los riesgos R01–R08 que detecte el análisis de riesgos (Tabla 2).

Escenario de aplicación	Descripción	Medidas relacionadas
Transversal	Medidas aplicables a cualquier sistema de IA según su finalidad prevista, su categoría, su contexto de utilización y su arquitectura.	[org.1] Política de seguridad [org.2] Normativa de seguridad [org.3] Procedimientos de seguridad [org.4] Proceso de autorización [op.pl.1] Análisis de riesgos [op.pl.2] Arquitectura de seguridad [op.pl.3] Adquisición de nuevos componentes [op.pl.4] Dimensionamiento / Gestión de la capacidad [op.acc.1] Identificación [op.acc.2] Requisitos de acceso [op.acc.4] Proceso de gestión de derechos de acceso [op.exp.1] Inventario de activos [op.exp.7] Gestión de incidentes [op.exp.8] Registro de la actividad [op.nub.1] Protección de servicios en la nube [op.mon.1] Detección de intrusión [mp.per.3] Concienciación [mp.per.4] Formación [mp.info.1] Datos personales [mp.info.6] Copia de seguridad

		[mp.s.2] Protección de servicios y aplicaciones web
Alto riesgo y/o sistema autónomo	Sistemas de IA de alto riesgo conforme a la definición detallada en el apartado 4.1 o sistemas de IA autónomos según la definición indicada en el apartado 2. Medidas adicionales a las transversales.	[op.acc.3] Segregación de funciones y tareas [op.exp.2] Configuración de seguridad [op.exp.3] Gestión de la configuración de seguridad [op.exp.5] Gestión de cambios [op.exp.8] Refuerzo de registro de la actividad [op.ext.1] Contratación y acuerdos de nivel de servicio [op.ext.3] Protección de la cadena de suministro [op.ext.4] Interconexión de sistemas [op.mon.3] Vigilancia [mp.sw.1] Desarrollo de aplicaciones
Riesgos específicos identificados	Aplicable a cualquier sistema de IA según los riesgos identificados en el análisis de riesgos. Medidas adicionales a las transversales.	Medidas seleccionadas según el riesgo (Tabla 2).

Tabla 1 Matriz de escenarios y medidas relacionadas

La siguiente tabla recoge las medidas asociadas a los riesgos R01–R08 definidos en el apartado 4.2. Los sistemas de IA pueden adoptar estas medidas para mitigar los riesgos identificados tras la realización del análisis de riesgos:

Código	Medida del ENS	Riesgos identificados
op	**Marco operacional**	
op.acc	**Control de acceso**	
op.acc.3	Segregación de funciones y tareas	R05
op.exp	**Explotación**	
op.exp.2	Configuración de seguridad	R01, R02, R03, R04, R05, R06
op.exp.3	Gestión de la configuración de seguridad	R01, R02, R03, R04, R05, R06
op.exp.5	Gestión de cambios	R08
op.exp.8	Refuerzo de registro de la actividad	R07
op.ext	**Recursos externos**	
op.ext.1	Contratación y acuerdos de nivel de servicio	R08
op.ext.3	Protección de la cadena de suministro	R08
op.ext.4	Interconexión de sistemas	R05
op.mon	**Monitorización del sistema**	
op.mon.3	Vigilancia	R03, R04, R05, R06, R07

mp	Medidas de protección	
mp.sw	**Protección de aplicaciones informáticas**	
mp.sw.1	Desarrollo de aplicaciones	R05

Tabla 2 Medidas adicionales para sistemas de IA por riesgo

7. GLOSARIO DE TÉRMINOS

El presente listado de terminología y definiciones recoge los principales conceptos utilizados a lo largo del documento, con el fin de establecer una base terminológica común.

- **Alfabetización en IA:** Medidas destinadas a apoyar la formación en materia de IA de su personal y demás personas que se encarguen en su nombre del funcionamiento y utilización de sistemas de IA, teniendo en cuenta sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso de los sistemas de IA y las personas o los colectivos de personas en que se van a utilizar dichos sistemas.
- **Alucinación:** Las alucinaciones se producen cuando los LLM llenan los vacíos en sus datos de entrenamiento utilizando patrones estadísticos, sin comprender realmente el contenido. Como resultado, el modelo puede producir respuestas que parecen correctas pero que son completamente infundadas.
- **Ciclo de vida de la IA:** Evolución de un sistema, producto, servicio, proyecto u otra entidad creada por el ser humano, desde su concepción hasta su retirada. Se divide en las siguientes fases:
  - Concepción
  - Diseño y desarrollo
  - Verificación y validación
  - Despliegue
  - Funcionamiento y supervisión
  - Reevaluación
  - Retirada
- **Conector:** Interconexión entre distintos componentes del sistema de IA que permite leer o ejecutar las acciones dentro de los permisos concedidos.
- **Contexto:** Es el flujo combinado de tokens que el modelo puede procesar simultáneamente en una sola interacción; incluye el system prompt, las entradas del usuario, los documentos recuperados, el historial de la conversación y el estado de memoria.

- **Entrenamiento:** Proceso de aprendizaje mediante el cual el sistema de IA adquiere conocimiento mediante algoritmos basados en datos de entrenamiento.
- **Espacio o piloto controlado de pruebas para IA:** Marco controlado que permite la posibilidad de desarrollar, entrenar, validar y probar, en condiciones reales con arreglo a un plan del espacio controlado de pruebas y durante un tiempo limitado, bajo supervisión regulatoria.
- **Finalidad prevista:** El uso concreto para el que el proveedor diseñó el sistema, tal como lo describe en manuales, fichas técnicas y materiales de venta. Define los límites de uso legítimo.
- **Fuente de conocimiento:** Es el repositorio de información corporativa (bases de datos vectoriales, archivos, APIs) a partir del cual se extrae información relevante para enriquecer el contexto del modelo antes de generar una respuesta.
- **Guardrail (guardarraíles):** Son mecanismos y filtros de protección independientes que actúan como barreras éticas, legales y operativas para analizar las entradas y salidas del modelo, garantizando que el sistema opere dentro de los límites y políticas de seguridad configurados.
- **Human-in-the-loop (HITL):** Es un principio de control y gobernanza que exige la intervención, revisión o aprobación explícita de una persona antes de que el sistema de IA ejecute decisiones o acciones de alto impacto, irreversibles, financieras, legales o sobre datos críticos.
- **Human-on-the-loop (HOTL):** El humano no interviene en la ejecución del sistema de IA, pero vigila el proceso general y puede detenerlo, corregirlo o intervenir solo si detecta una anomalía o un error. La supervisión humana en este caso se denomina ex post, e implica que la acción está totalmente automatizada, aunque se deba habilitar posteriormente la posibilidad de que una intervención humana permita revertirla.
- **Inferencia:** Es la fase en la que un modelo de IA procesa la información de entrada y deduce o genera un resultado de salida, tal como predicciones, contenidos sintéticos, recomendaciones o decisiones.
- **Jailbreak:** modalidad específica de inyección de prompt en la que un usuario o atacante proporciona entradas diseñadas deliberadamente para eludir, neutralizar o saltarse los controles de seguridad, las políticas de uso y los protocolos de restricción establecidos en el modelo base.
- **Memoria:** Es el estado o contexto retenido por un sistema de IA entre interacciones, pudiendo ser efímero/transitorio (Short-Term, acotado a la sesión activa) o persistente (Long-Term, almacenado a lo largo del tiempo en bases de datos o servicios de memoria para influir en futuras interacciones).
- **Modelo de IA:** Representación lógica de un proceso o sistema, obtenida mediante aprendizaje automático. A diferencia del software tradicional, su valor no está en el código sino en el conocimiento adquirido durante el

entrenamiento. Requiere protección frente a ingeniería inversa y robo de modelos.

- **Modelo de IA de uso general:** Modelo entrenado a gran escala con autosupervisión, capaz de realizar tareas muy diversas sin haber sido diseñado para ninguna en particular. Puede integrarse en múltiples aplicaciones. Quedan fuera los modelos usados solo en fase de investigación o prototipado, antes de salir al mercado.
- **Modificación sustancial:** Cambio en un sistema de IA, producido tras su introducción en el mercado o puesta en servicio, que no haya sido previsto, o que dé lugar a una modificación de la finalidad prevista para la que dicho sistema fue evaluado.
- **Parámetros del sistema de IA:** Son variables de configuración que son internas del sistema de IA y cuyos valores pueden estimarse a partir de los datos.
- **Prompt:** Es el contenido de texto, instrucción, pregunta o ejemplo de datos introducido como entrada (input) en un modelo de lenguaje o agente para guiar su respuesta o desencadenar una acción.
- **Proveedor de IA:** Quien desarrolla un sistema o modelo de IA y lo lanza al mercado bajo su propio nombre, sea cobrando por ello o gratuitamente.
- **Responsable del despliegue:** Quien usa el sistema de IA en su operativa real, bajo su propia responsabilidad. Queda excluido el uso estrictamente personal y no profesional.

  Riesgo sistémico: El RIA introduce una categoría especial para los "modelos de IA de uso general" clasificados con riesgo sistémico (riesgos graves de perturbación de sectores críticos, procesos democráticos o difusión de contenidos ilícitos a gran escala) si:

  - La Comisión decide que tienen capacidades de gran impacto evaluadas con metodologías y parámetros de referencia técnicos, basándose en los criterios establecidos en el anexo XIII del RIA.
  - Si la cantidad acumulada de cálculo utilizada para su entrenamiento es superior a un número determinado de operaciones de coma flotante (FLOP).
- **Sesgo:** Desviación sistemática respecto a un estándar de referencia. Puede tener origen estadístico (muestra de datos no representativa) o humano (prejuicios introducidos durante el diseño o etiquetado).
- **Sistema de IA:** Máquina que procesa entradas para generar salidas (predicciones, decisiones, recomendaciones) con cierto grado de autonomía. No ejecuta instrucciones fijas: infiere cómo actuar a partir de los datos que recibe y puede ajustar su comportamiento después de haber sido desplegado.

- **Solidez:** Capacidad del sistema para mantener su rendimiento bajo condiciones adversas: datos ruidosos, entradas no previstas en el entrenamiento, o ataques deliberados.
- **Supervisión humana:** Requisito aplicable a los sistemas de IA consistente en que dichos sistemas sean diseñados y desarrollados de forma que, durante su periodo de uso, puedan ser supervisados eficazmente por personas físicas.
- **System prompt:** Es un bloque de instrucciones privilegiado y generalmente no visible para el usuario final que establece el comportamiento general, el tono, las reglas de rol, los límites de capacidad y las directrices de seguridad que el modelo debe acatar a lo largo de la interacción.
- **Transparencia:** Disponibilidad de información sobre el sistema en su conjunto: datos de entrenamiento, limitaciones conocidas, alcance y propósito. Complementa a la explicabilidad (que mira el resultado concreto) mirando el sistema completo.

8. MARCO NORMATIVO Y DE REFERENCIA

La incorporación de los sistemas de IA a los servicios y sistemas regulados por el ENS debe basarse en los principios de seguridad establecidos dentro de un marco normativo de referencia. Entre ellos:

[Ref.1] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial actualizado por el por el Reglamento (UE) 2026/1744, de 8 de julio de 2026 (Ómnibus digital sobre IA).

[Ref.2] Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales (RGPD).

[Ref.3] Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un nivel común y elevado de ciberseguridad en toda la Unión (NIS2).

[Ref.4] Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD).

[Ref.5] Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad (ENS).

[Ref.6] Guía CCN-STIC 801 Responsabilidades y funciones

[Ref.7] Guía CCN-STIC 861. Criterios de Aplicabilidad de las medidas del ENS frente a la IA ofensiva

[Ref.8] Guía CCN-CERT BP/36 de buenas prácticas frente al modelo de IA ofensiva

[Ref.9] OWASP Top 10 for LLM Applications 2026.

[Ref.10] Guía 02 de definiciones AESIA.

[Ref.11] Guía 11 de ciberseguridad AESIA.

9. ACRÓNIMOS

AESIA	Agencia Española de Supervisión de la Inteligencia Artificial
API	Interfaz de Programación de Aplicaciones (Application Programming Interface)
CCN	Centro Criptológico Nacional
ENS	Real Decreto 311/2022, de 3 de mayo, por el que se regula el Esquema Nacional de Seguridad
FLOP	Floating Point Operations
GPU	Graphics Processing Unit
HITL	Human-in-the-loop
HOTL	Human-on-the-loop
IA	Inteligencia Artificial
LLM	Large Language Model
LOPDGDD	Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales
NIS2	Directiva (UE) 2022/2555, relativa a medidas destinadas a garantizar un nivel común y elevado de ciberseguridad en toda la Unión
NPU	Neural Processing Unit
OWASP	Open Worldwide Application Security Project
PSI	Política de Seguridad de la Información
RGPD	Reglamento (UE) 2016/679, General de Protección de Datos
RIA	Reglamento (UE) 2024/1689, por el que se establecen normas armonizadas en materia de inteligencia artificial.
SaaS	Software as a Service
UE	Unión Europea

Anexo A CASO DE USO: Asistente normativo

Se presenta a continuación un ejemplo de aplicación de la guía sobre un sistema de IA que no está clasificado como alto riesgo ni se considera un sistema de IA que ejecute acciones de forma autónoma. El ejemplo está enfocado a un asistente conversacional que en este caso concreto actúa como asistente normativo que facilita la consulta de documentación interna.

El objetivo de este apartado es trasladar los conocimientos adquiridos a casos reales que se pueden dar en una entidad y completar el proceso de uso seguro de los sistemas de IA.

Para ello, se van desarrollando cada uno de los puntos del proceso (arquitectura, inventario de activos, análisis de riesgos, etc.) como apoyo a las entidades que cuenten con un sistema de IA al que le apliquen medidas en el escenario transversal.

Asimismo, se estudia la posibilidad definida en el apartado **¡Error! No se encuentra el origen de la referencia.** de incluir medidas adicionales para mitigar los riesgos específicos detectados en el análisis de riesgos.

Este caso de uso no pretende englobar todos los sistemas de IA del escenario transversal, sino orientar a las entidades sobre cómo seguir el proceso de securización con sus propios recursos.

A1. Descripción del caso de uso

Una entidad usa un asistente normativo basado en IA para facilitar la consulta de documentación interna (normativa, procedimientos, manuales, guías técnicas…) autorizada y preparada por la entidad. Adicionalmente, el sistema de IA tiene acceso a documentación pública necesaria para el funcionamiento del asistente.

El sistema permite a los usuarios del asistente formular consultas en lenguaje natural y obtener respuestas generadas a partir de las fuentes de información incorporadas al sistema, con el objetivo de facilitar la localización, comprensión y utilización de la documentación disponible.

La finalidad prevista del sistema es servir como herramienta de apoyo en la consulta de normativa.

Por tanto, el sistema:

- Opera bajo supervisión humana, siendo los usuarios los responsables de revisar e interpretar los resultados obtenidos. No toma decisiones de forma autónoma.
- No accede a fuentes de información distintas a las autorizadas por la entidad.
- No ejecuta acciones sobre otras aplicaciones, sistemas o servicios.

A2. Arquitectura del sistema de IA

La arquitectura del sistema de IA es la siguiente:



Ilustración 2 Arquitectura de sistema de IA en el caso de uso asistente normativo

Tal y como se observa en la ilustración, todos los componentes del sistema de IA estarían alojados en la entidad a excepción del modelo, que estaría prestado por un tercero. Cada componente cumple la siguiente función:

- **Usuario:** Origen de la consulta, petición o instrucción al sistema de IA, como puede ser una persona o una aplicación.
- **Filtros de entrada:** Mecanismos que validan, normalizan y analizan la solicitud recibida, aplicando políticas de seguridad, clasificación o restricciones de uso.
- **Motor de búsqueda:** Componente que localiza y recupera la información más relevante desde las fuentes autorizadas para responder a la consulta.
- **Fuentes autorizadas:** Repositorios de información autorizados por los responsables de la entidad a los que el sistema de IA tenga acceso, como pueden ser la documentación interna, bases de datos, aplicaciones corporativas o fuentes externas autorizadas.
- **LLM externo:** Modelo de IA externo escogido por la entidad, encargado de procesar la consulta y el contexto recuperado para generar una respuesta coherente y útil. Este modelo es operado y gestionado por el tercero desde sus servidores.
- **Filtros de salida:** Controles aplicados sobre la respuesta generada para verificar aspectos de seguridad, privacidad, calidad, formato y cumplimiento de políticas.
- **Respuesta:** Resultado final entregado al usuario, que puede incluir texto, referencias, recomendaciones, advertencias, explicaciones o decisiones.

A3. Inventario de activos

A partir de la arquitectura definida, podemos distinguir los siguientes activos, agrupados conforme a lo definido en el apartado 3:

- **Activos de datos:** Únicamente se incorporarán documentos previamente seleccionados, revisados y autorizados por la entidad para formar parte de las fuentes de información del asistente:
  - Documentación normativa pública incorporada al sistema.
  - Procedimientos internos de la entidad.
  - Manuales.
  - Guías.
  - Consultas realizadas por los usuarios.
  - Respuestas generadas por el sistema.
- **Activos de software, aplicación y modelos:**
  - Portal web o interfaz de usuario.
  - Filtros de entrada y de salida.
  - Motor de búsqueda.
  - Base de datos o repositorio con las fuentes de conocimiento.
  - Prompt del sistema (system prompt)
  - Parámetros de configuración.
- **Activos de infraestructura y ejecución:**
  - Servidor o servicio donde se aloja la interfaz de usuario.
  - Infraestructura de ejecución del motor de búsqueda.
  - Infraestructura de almacenamiento de las fuentes de conocimiento.
  - Servicios de autenticación corporativa.
- **Activos de supervisión y trazabilidad:**
  - Registros de acceso.
  - Registros de consultas realizadas.
  - Registros de búsquedas y documentos recuperados.
  - Registros de respuestas generadas.
  - Registros de errores y eventos de seguridad.
  - Registros de auditoría.
  - Mecanismos de supervisión humana.
- **Activos externos en la cadena de valor (terceros):**

- Modelo de IA (LLM externo), sus versiones y la infraestructura que lo soporta.

A4. Tipo de sistema de IA

Conforme a la clasificación de riesgos definida por el RIA [Ref.1] y detallada en el apartado 4.1 y atendiendo a su finalidad prevista, arquitectura y caso de uso, el sistema no se encuentra comprendido en ninguno de los ámbitos de alto riesgo definidos por el RIA, no realiza prácticas prohibidas y el modelo del sistema de IA no es de uso general con riesgo sistémico.

Por tanto, se considera como un **sistema de IA al que se le aplican las obligaciones de transparencia (para el proveedor del modelo y la entidad como responsable del despliegue)** debido a su interacción con humanos.

A5. Riesgos específicos identificados

- **Divulgación y tratamiento inadecuado de información (R01)**

Existe el riesgo de que un usuario introduzca información sensible en el sistema o de que determinadas respuestas revelen información no autorizada durante su utilización.

La materialización de este riesgo afectaría principalmente a la confidencialidad de la información gestionada por la entidad.

- **Acceso indebido a la información (R02)**

El sistema podría recuperar documentos cuyo acceso no corresponde al usuario que realiza la consulta. Este riesgo presenta especial relevancia en la arquitectura propuesta, al utilizar un modelo externo conectado a un repositorio documental.

Este riesgo afecta principalmente a la confidencialidad de la información y justifica la necesidad de controlar los permisos asociados tanto al repositorio como a los componentes del sistema de IA.

- **Resultados o decisiones incorrectas o no fiables (R03)**

Aunque el sistema utilice exclusivamente documentación autorizada por la entidad, el modelo puede generar respuestas incorrectas, incompletas, desactualizadas o insuficientemente fundamentadas debido a las limitaciones inherentes de los sistemas de IA.

Este riesgo resulta especialmente relevante en un asistente, ya que los usuarios pueden asumir incorrectamente que las respuestas generadas son siempre correctas o constituyen una interpretación válida de la normativa aplicable, afectando directamente a la integridad.

- **Manipulación de instrucciones, contexto y fuentes de conocimiento (R04)**

Un usuario podría intentar alterar el comportamiento previsto mediante instrucciones maliciosas o contenido manipulado incluido en la documentación consultada.

- **Falta de supervisión, trazabilidad y control organizativo (R07)**

La utilización del sistema requiere disponer de mecanismos que permitan conocer quién realizó una consulta, qué información fue utilizada y qué respuesta fue generada, facilitando así la supervisión, auditoría e investigación de incidencias.

- **Dependencia de terceros y cadena de suministro (R08)**

Este caso de uso depende del modelo proporcionado por un tercero, por lo que la entidad debe contar con medidas para asegurar la cadena de suministro protegiendo la continuidad del servicio y la transparencia del funcionamiento del componente prestado por el tercero.

A6. Aplicación de medidas de seguridad

A partir de la consideración del sistema de IA y de los riesgos específicos identificados, se determinan las medidas de seguridad aplicables conforme al apartado **¡Error! No se encuentra el origen de la referencia.** de la presente guía.

Al tratarse de un sistema de IA que no es de alto riesgo ni tiene capacidades de actuación autónoma, resultan de aplicación las medidas de carácter general incluidas en el escenario **transversal.**

Adicionalmente, como se explica en el apartado **¡Error! No se encuentra el origen de la referencia.**, la entidad podría incorporar medidas adicionales cuando le aplique un riesgo específico identificado en el apartado anterior.

La siguiente tabla indica las medidas a tener en cuenta para este caso de uso:

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
org	**Marco organizativo**				
org.1	Política de seguridad	X		Entidad	
org.2	Normativa de seguridad	X		Entidad	
org.3	Procedimientos de seguridad	X		Entidad	
org.4	Proceso de autorización	X		Entidad	
op	**Marco operacional**				
op.pl					
op.pl.1	Análisis de riesgos	X		Compartida	La entidad tendrá en cuenta durante el proceso de análisis y gestión de riesgos adicionalmente aquellos activos, amenazas y riesgos derivados del uso del sistema de IA. El tercero deberá identificar, describir y compartir aquellos riesgos que afecten al modelo de IA proporcionado a la entidad y que deban incorporarse al análisis de riesgos.
op.pl.2	Arquitectura de seguridad	X		Entidad	

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
op.pl.3	Adquisición de nuevos componentes	X		Compartida	La entidad considerará los componentes del sistema de IA durante el proceso de adquisición. El tercero identificará y facilitará información acerca del modelo y sus versiones a la entidad.
op.pl.4	Dimensionamiento / Gestión de la capacidad	X		Compartida	La entidad tendrá en cuenta los requisitos de computación, almacenamiento, personal y comunicaciones específicos de los sistemas de IA. El tercero indicará a la entidad los requisitos y límites necesarios para operar el modelo de forma segura.
op.acc	**Control de acceso**				
op.acc.1	Identificación	X		Compartida	La entidad tendrá en cuenta las identidades de los diferentes componentes de los sistemas de IA, permitiendo su identificación y trazabilidad. El tercero explicará a la entidad cómo identificar de forma única al modelo y sus versiones.

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
op.acc.2	Requisitos de acceso	X		Compartida	La entidad asignará los permisos estrictamente necesarios en los componentes del sistema de IA para el desempeño de sus funciones. El tercero facilitará información sobre cómo aplicar los permisos sobre el modelo y sus versiones y qué opciones de configuración ofrece a la entidad.
op.acc.4	Proceso de gestión de derechos de acceso	X		Compartida	La entidad trasladará el principio de mínimo privilegio y necesidad de conocer a los sistemas de IA. El tercero establecerá las pautas para gestionar el control de acceso y dará las instrucciones para que la entidad que consuma el modelo pueda gestionarlo de forma adecuada.
op.exp	**Explotación**				
op.exp.1	Inventario de activos	X		Compartida	La entidad identificará debidamente en el inventario de activos todos los componentes del sistema de IA.

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
					El tercero proporcionará a la entidad un listado completo del modelo y sus versiones.
op.exp.2	Configuración de seguridad		R01, R02, R03, R04	Compartida	La entidad configurará los sistemas de IA implementando medidas técnicas, operacionales y en tiempo de ejecución que validen, filtren y acoten las entradas, procesos y salidas del sistema de IA para mantener su comportamiento seguro, conforme y dentro del inventario o caso de uso autorizado. El tercero deberá garantizar el aislamiento del modelo y la restricción de permisos.
op.exp.3	Gestión de la configuración de seguridad		R01, R02, R03, R04	Entidad	
op.exp.5	Gestión de cambios		R08	Compartida	La entidad desarrollará el proceso de gestión de cambios relativo al sistema de IA haciendo especial énfasis en los cambios de los componentes del sistema de IA (como modelos) que puedan

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
					tener repercusiones en la operación y calidad del sistema. El tercero informará a la entidad de los cambios relevantes en el modelo.
op.exp.7	Gestión de incidentes	X		Compartida	La entidad contará con los recursos necesarios para analizar, notificar, contener y resolver incidentes de seguridad que afecten al sistema de IA. El tercero deberá implementar y definir los mecanismos para detectar comportamientos anómalos del modelo con el fin de prevenir y detectar posibles incidentes.
op.exp.8	Registro de la actividad	X		Entidad	
op.exp.8	Refuerzo de registro de la actividad		R07	Entidad	
op.ext	**Recursos externos**				
op.ext.1	Contratación y acuerdos de nivel de servicio		R08	Tercero	El tercero se comprometerá a cumplir los requisitos de transparencia y de colaboración en la gestión de seguridad del modelo impuestos por la entidad

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
op.ext.3	Protección de la cadena de suministro		R08	Compartida	La entidad deberá analizar el impacto que puede tener sobre su sistema un incidente accidental o deliberado que tenga su origen en un proveedor del sistema de IA y esta dependencia se tendrá en cuenta en el plan de continuidad de la entidad. A su vez el tercero cumplirá las obligaciones como proveedor impuestas en el RIA para garantizar la seguridad del sistema de IA.
op.nub	**Servicios en la nube**				
op.nub.1	Protección de servicios en la nube	X		Tercero	El tercero deberá ser conforme con el ENS o regirse por las guías CCN-STIC relativas a la protección de servicios de IA en la nube.
op.mon	**Monitorización del sistema**				
op.mon.1	Detección de intrusión	X		Entidad	
op.mon.3	Vigilancia		R03, R04, R07	Compartida	La entidad llevará un registro de la configuración de seguridad y los usuarios activos, volúmenes, fuentes consultadas, errores, bloqueos, consumo,

Código	Medida del ENS	Transversal	Adicional por riesgos específicos identificados	Responsable de implantación	Criterio de implantación
					disponibilidad, conexiones y eventos anómalos. Asimismo, el tercero debe garantizar la configuración de indicadores para detectar desviaciones, crecimiento no previsto, usos residuales o dependencia creciente, así como indicadores que permitan a la entidad activar la intervención humana.
mp	**Medidas de protección**				
mp.per	**Gestión del personal**				
mp.per.3	Concienciación	X		Entidad	
mp.per.4	Formación	X		Entidad	
mp.info	**Protección de la información**				
mp.info.1	Datos personales	X		Entidad	
mp.info.6	Copia de seguridad	X		Entidad	
mp.s	**Protección de los servicios**				
mp.s.2	Protección de servicios y aplicaciones web	X		Entidad	

Tabla 3 Medidas caso de uso asistente

Anexo B CASO DE USO: Sistema autónomo de evaluación de competencias

Se presenta a continuación un ejemplo de aplicación de la guía sobre un sistema de IA clasificado como alto riesgo conforme a la definición detallada en el apartado 4.1 y/o que ejecuta acciones de forma autónoma. El ejemplo está enfocado a un sistema de IA para automatización de tareas que en este caso concreto actúa como un sistema autónomo de evaluación de competencias.
El objetivo de este apartado es trasladar los conocimientos adquiridos a casos reales que se pueden dar en una entidad y completar el proceso de uso seguro de los sistemas de IA.
Para ello, se van desarrollando cada uno de los puntos del proceso (arquitectura, inventario de activos, análisis de riesgos, etc.) como apoyo a las entidades que cuenten con un sistema de IA al que le apliquen medidas del escenario Alto riesgo y/o sistema autónomo.
Este caso de uso no pretende englobar todos los sistemas de IA del escenario Alto riesgo y/o sistema autónomo, sino orientar a las entidades sobre cómo seguir el proceso de securización con sus propios recursos.

B1. Descripción del caso de uso

Una entidad contrata un sistema de IA autónomo a un tercero para generar y evaluar pruebas para valorar las competencias de los empleados.

El sistema de IA tendrá acceso a un repositorio donde se almacenará documentación autorizada y preparada por la entidad sobre la que se definen las pruebas. Se contará con un primer agente que genere exámenes para los empleados.

Adicionalmente, existirá una plataforma de enseñanza donde los usuarios subirán las pruebas finalizadas, las cuales serán corregidas de manera automática por un segundo agente que también tiene acceso al repositorio. Una vez las pruebas sean corregidas, el sistema de IA publicará las notas resultantes en la plataforma de la entidad estableciendo si el candidato es apto o no para la promoción.

En caso de que se detecte alguna anomalía en el proceso, se activan las alertas de supervisión humana.

La finalidad prevista del sistema es servir como herramienta de generación, corrección y calificación automatizada de pruebas destinadas a evaluar las competencias de los empleados.

Por tanto, el sistema:

- Opera de manera autónoma, generando, corrigiendo y evaluando las pruebas.
- Ejecuta acciones sobre otras aplicaciones, sistemas o servicios de la entidad. En este caso, publica las notas resultantes en la plataforma de la entidad.
- No accede a fuentes de información distintas a las autorizadas por la entidad.

B2. Arquitectura del sistema de IA

La arquitectura del sistema de IA es la siguiente:

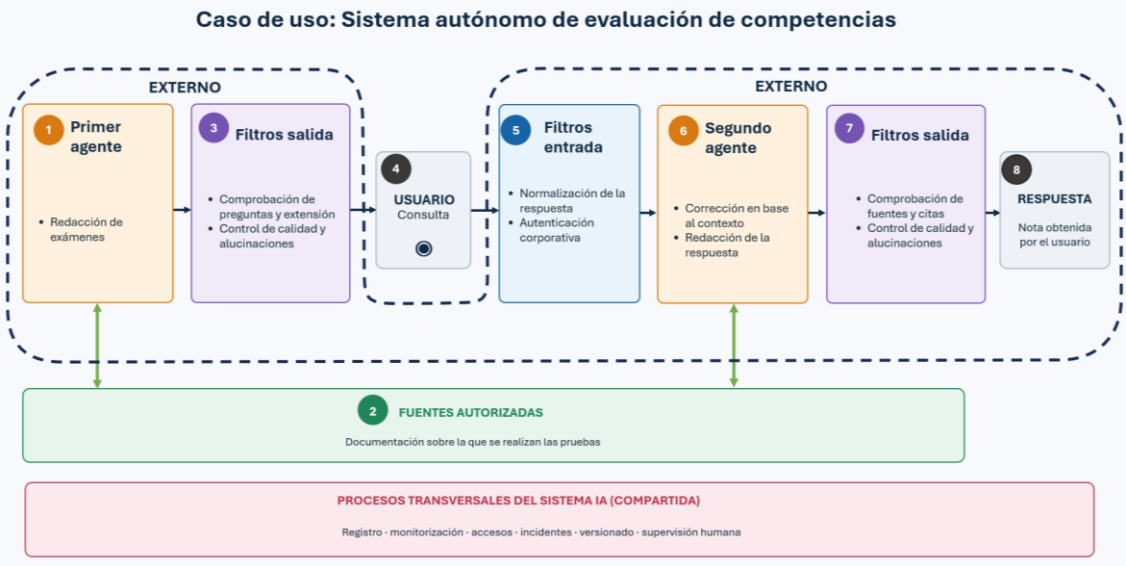

Ilustración 3 Arquitectura de sistema de IA en el caso de uso sistema autónomo de evaluación de competencias

Tal y como se observa en la ilustración, todos los componentes del sistema de IA estarían alojados en el tercero salvo la interacción del usuario que estaría en la entidad. Cada componente cumple la siguiente función:

- **Primer agente:** Elabora las pruebas en base a la documentación proporcionada por la entidad. Incluye el LLM.
- **Fuentes autorizadas:** Documentación sobre la que la entidad quiere evaluar al usuario.
- **Filtros de salida:** Controles aplicados sobre el examen generado para verificar los aspectos del formato y cumplimiento de políticas.
- **Usuario:** Completa la prueba generada por el primer agente y la vuelve a introducir en el sistema de IA.
- **Filtros de entrada:** Mecanismos que validan, normalizan y analizan la prueba recibida, aplicando políticas de seguridad, clasificación o restricciones de uso.
- **Segundo agente:** Revisa y corrige la prueba del usuario en base a la documentación proporcionada por la entidad y devuelve la nota obtenida. Incluye el LLM.
- **Filtros de salida:** Controles aplicados sobre la respuesta generada para verificar aspectos de seguridad, privacidad, calidad, formato y cumplimiento de políticas.
- **Respuesta:** Nota final del usuario publicada en la plataforma de la entidad.

B3. Inventario de activos

A partir de la arquitectura definida, podemos distinguir los siguientes activos, agrupados conforme a lo definido en el apartado 3:

- **Activos de datos:** Únicamente se incorporarán documentos previamente seleccionados, revisados y autorizados por la entidad para formar parte de las fuentes de información del asistente:
  - Documentación incorporada al sistema.
  - Pruebas realizadas por los usuarios.
- **Activos de software, aplicación y modelos:**
  - Base de datos o repositorio con las fuentes de conocimiento.
- **Activos de infraestructura y ejecución:**
  - Servidor o servicio donde se aloja la interfaz de usuario.
  - Infraestructura de almacenamiento de las fuentes de conocimiento.
  - Infraestructura donde se aloja y ejecuta el modelo de IA.
  - Servicios de autenticación corporativa.
- **Activos de supervisión y trazabilidad:**
  - Registros de acceso.
  - Registros de pruebas realizadas.
  - Registros de búsquedas y documentos recuperados.
  - Registros de notas generadas.
- **Activos externos en la cadena de valor (terceros):**
  - Agentes externos, incluyendo sus modelos y sus versiones, así como los motores de búsqueda.
  - Portal web o interfaz de usuario.
  - Filtros de entrada y de salida.
  - Parámetros de configuración.
  - Registros de errores y eventos de seguridad.
  - Registros de auditoría.
  - Mecanismos de supervisión humana.

B4. Tipo de sistema de IA

Conforme a la clasificación de riesgos definida por el RIA [Ref.1] y detallada en el apartado 4.1, y atendiendo a su finalidad prevista, arquitectura y caso de uso, es un **sistema de IA de alto riesgo** comprendido en la finalidad del anexo III, punto 4.b) del

RIA, “Sistemas de IA destinados a ser utilizados para tomar decisiones que afecten a las condiciones de las relaciones de índole laboral o a la promoción o rescisión de relaciones contractuales de índole laboral, para la asignación de tareas a partir de comportamientos individuales o rasgos o características personales o para supervisar y evaluar el rendimiento y el comportamiento de las personas en el marco de dichas relaciones”.

Adicionalmente, se trata de un **sistema de IA autónomo**, según la definición del apartado 2, ya que opera la mayoría del tiempo bajo su propio criterio sin que sea necesaria la intervención humana.

B5. Riesgos específicos identificados

- **Divulgación y tratamiento inadecuado de información (R01)**

El sistema de IA trata información relativa a procesos de evaluación, resultados obtenidos por los usuarios y datos que permiten identificarlos. Además, dispone de acceso a la documentación utilizada para la generación y corrección de las pruebas.

La divulgación no autorizada de dicha información podría afectar a la confidencialidad de los procesos de evaluación y exponer información personal o profesional de los usuarios evaluados.

- **Acceso indebido a la información (R02)**

Una configuración incorrecta o accesos no autorizados podrían permitir que usuarios o sistemas accedieran a exámenes, preguntas, respuestas, criterios de evaluación o resultados correspondientes a otros usuarios.

Esta situación podría comprometer la confidencialidad e integridad de la información utilizada durante el proceso de promoción, afectando a la fiabilidad y legitimidad de los resultados obtenidos.

- **Resultados o decisiones incorrectas o no fiables (R03)**

El sistema podría generar preguntas inadecuadas, ambiguas, incompletas o realizar correcciones erróneas de las respuestas aportadas por los usuarios, produciendo calificaciones incorrectas que afecten al proceso de promoción y a las decisiones adoptadas por la entidad sobre los usuarios evaluados.

Este riesgo adquiere especial relevancia al tratarse de un sistema cuyos resultados influyen directamente en la valoración de personas.

- **Manipulación de instrucciones, contexto y fuentes de conocimiento (R04)**

Un usuario podría intentar incorporar instrucciones o respuestas diseñadas para modificar el comportamiento esperado del sistema durante la generación de pruebas o durante su corrección.

Además, modificaciones no autorizadas de la documentación utilizada como fuente de conocimiento podrían alterar los criterios aplicados por los agentes, comprometiendo

la integridad del proceso de evaluación y afectando a la fiabilidad de las calificaciones obtenidas.

- **Uso indebido de conectores, herramientas y capacidades de actuación (R05)**

El sistema de IA dispone de capacidades para interactuar con otros sistemas, publicando las calificaciones obtenidas en la plataforma de enseñanza de la entidad. Por tanto, una configuración incorrecta, fallos en los agentes o comportamientos no previstos, podrían provocar la publicación de resultados erróneos, la modificación de calificaciones o la asignación incorrecta de resultados a usuarios distintos de los evaluados.

Al tratarse de un sistema que realiza acciones de forma autónoma sobre otros sistemas, este riesgo adquiere especial relevancia dentro del caso de uso analizado.

- **Resultados sesgados o no representativos (R06)**

El sistema podría generar evaluaciones inconsistentes, no objetivas o insuficientemente representativas debido a limitaciones del modelo, deficiencias en los criterios de evaluación o problemas existentes en la documentación de referencia. Por otro lado, determinadas características de las preguntas generadas podrían favorecer o perjudicar de forma no intencionada a determinados usuarios respecto a otros.

Como consecuencia, podrían producirse valoraciones que no reflejan adecuadamente las competencias reales de los participantes, afectando a la equidad y objetividad del proceso de promoción.

- **Falta de supervisión, trazabilidad y control organizativo (R07)**

La utilización del sistema requiere disponer de mecanismos que permitan conocer qué examen fue generado, qué documentación fue utilizada para su elaboración, qué respuestas fueron evaluadas, qué criterios fueron aplicados durante la corrección y qué calificación fue finalmente publicada.

La ausencia de estos registros o de mecanismos de supervisión adecuados dificultaría la trazabilidad del proceso, la revisión de reclamaciones o la justificación de los resultados obtenidos.

Al tratarse de un sistema de IA de alto riesgo y con un cierto grado de autonomía, la trazabilidad de las actuaciones realizadas por el sistema resulta especialmente relevante.

- **Dependencia de terceros y cadena de suministro (R08)**

Este caso de uso depende de un modelo de IA proporcionado por un tercero y de los componentes asociados necesarios para su funcionamiento. Cualquier modificación no controlada, indisponibilidad del servicio, vulnerabilidad o cambio en las funcionalidades ofrecidas por el tercero podría afectar al correcto funcionamiento del sistema.

La entidad debe disponer de mecanismos que permitan gestionar adecuadamente la relación con el tercero, evaluar los riesgos asociados a la cadena de valor del sistema de IA y garantizar la continuidad, seguridad y trazabilidad del servicio prestado.

B6. Aplicación de medidas de seguridad

A partir de la consideración del sistema de IA y de los riesgos específicos identificados, se determinan las medidas de seguridad aplicables conforme al apartado **¡Error! No se encuentra el origen de la referencia.** de la presente guía.

Al tratarse de un sistema de IA de alto riesgo y autónomo, resultan de aplicación las medidas aplicables en el escenario **Alto riesgo y/o sistema autónomo**:

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
org	**Marco organizativo**				
org.1	Política de seguridad	X		Entidad	
org.2	Normativa de seguridad	X		Entidad	
org.3	Procedimientos de seguridad	X		Compartida	La entidad supervisará y coordinará los procedimientos de seguridad con el tercero. El tercero dispondrá de una serie de procedimientos que detallen de forma clara y precisa cómo operar los elementos del sistema de IA y los deberá comunicar a la entidad.
org.4	Proceso de autorización	X		Compartida	La entidad supervisará y coordinará el proceso de autorización con el tercero. El tercero dispondrá de un proceso formal de autorizaciones que cubra todos los elementos del sistema de IA.
op	**Marco operacional**				
op.pl					
op.pl.1	Análisis de riesgos	X		Compartida	La entidad tendrá en cuenta durante el proceso de análisis y gestión de riesgos adicionalmente

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					aquellos activos, amenazas y riesgos derivados del uso del sistema de IA. El tercero deberá identificar, describir y compartir aquellos riesgos que afecten a todos los componentes del sistema de IA proporcionado a la entidad y que deban incorporarse al análisis de riesgos.
op.pl.2	Arquitectura de seguridad	X		Compartida	La entidad supervisará y coordinará la arquitectura de sistema de IA. El tercero documentará y comunicará la arquitectura de seguridad del sistema de IA, documentando su funcionalidad, activos y relaciones.
op.pl.3	Adquisición de nuevos componentes	X		Compartida	La entidad supervisará y coordinará el proceso de adquisición de nuevos componentes del sistema de IA. El tercero identificará y facilitará información acerca de la adquisición de los componentes del sistema de IA a la entidad.

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
op.pl.4	Dimensionamiento / Gestión de la capacidad	X		Compartida	La entidad supervisará y coordinará los requisitos de computación, almacenamiento, personal y comunicaciones específicos del sistema de IA. El tercero trasladará a la entidad los requisitos y límites necesarios para operar el sistema de IA de forma segura.
op.acc	**Control de acceso**				
op.acc.1	Identificación	X		Compartida	La entidad supervisará y coordinará el proceso de identificación de los componentes del sistema. El tercero tendrá en cuenta las identidades de los diferentes componentes del sistema de IA (modelos, agentes…), permitiendo su identificación y trazabilidad.
op.acc.2	Requisitos de acceso	X		Compartida	La entidad asignará los permisos estrictamente necesarios en los componentes del sistema de IA para el desempeño de sus funciones. El tercero facilitará información sobre cómo aplicar los permisos

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					sobre el sistema de IA y sus versiones y qué opciones de configuración ofrece a la entidad.
op.acc.3	Segregación de funciones y tareas		X	Compartida	La entidad trasladará sus requisitos relativos a la segregación de funciones a los sistemas de IA cuando estos realicen tareas críticas como desarrollo, operación, autorización, etc. El tercero establecerá y configurará las pautas de segregación de funciones y dará las instrucciones necesarias para que la entidad que consuma el sistema de IA pueda gestionarlo de forma adecuada.
op.acc.4	Proceso de gestión de derechos de acceso	X		Compartida	El tercero proporcionará a la entidad información sobre los derechos de acceso para que la entidad pueda revisarlos y garantizar el principio de mínimo privilegio y necesidad de conocer.
op.exp	**Explotación**				
op.exp.1	Inventario de activos	X		Compartida	La entidad identificará debidamente en el inventario de

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					activos todos los componentes del sistema de IA. El tercero proporcionará a la entidad un listado completo de los componentes del sistema de IA.
op.exp.2	Configuración de seguridad		X	Compartida	La entidad supervisará y trasladará sus requisitos a nivel de implementación de medidas al tercero. El tercero garantizará la configuración de los sistemas de IA implementando medidas técnicas, operacionales y en tiempo de ejecución que validen, filtren y acoten las entradas, procesos y salidas del sistema de IA para mantener su comportamiento seguro, conforme y dentro del inventario o caso de uso autorizado.
op.exp.3	Gestión de la configuración de seguridad		X	Compartida	La entidad supervisará el mantenimiento de la configuración de la seguridad. El tercero mantendrá y comunicará a la entidad las

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					medidas de configuración de seguridad aplicadas.
op.exp.5	Gestión de cambios		X	Compartida	La entidad desarrollará el proceso de gestión de cambios relativo al sistema de IA haciendo especial énfasis en los cambios de los componentes del sistema de IA, como por ejemplo modelos, que puedan tener repercusiones en la operación y calidad del sistema. El tercero informará a la entidad de los cambios relevantes en los componentes del sistema de IA.
op.exp.7	Gestión de incidentes	X		Compartida	La entidad supervisará el procedimiento de gestión de incidentes del tercero y lo integrará con los de la entidad. El tercero deberá implementar, definir y notificar a la entidad un procedimiento de gestión de incidentes de seguridad (detección, comunicación, resolución, aprendizaje, etc.).
op.exp.8	Registro de la actividad	X		Compartida	La entidad transmitirá al tercero las necesidades relativas a los registros de auditoría dentro del sistema de IA e incorporará

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					registros que almacenen las interacciones del usuario con el sistema de IA. El tercero deberá garantizar las capacidades de registros de auditoría en los procesos internos del sistema de IA.
op.exp.8	Refuerzo de registro de la actividad		X	Compartida	Para reforzar el principio de transparencia, la entidad trasladará los requisitos relativos a registros específicos de actividades de cada uno de los componentes del sistema de IA. Por ejemplo, trasladará requisitos relativos al registro de errores en la formulación o corrección de las pruebas de evaluación de competencias. El tercero garantizará registros de monitorización de las actividades de cada uno de los componentes del sistema de IA.
op.ext	**Recursos externos**				
op.ext.1	Contratación y acuerdos de nivel de servicio		X	Compartida	La entidad trasladará en la contratación los requisitos del sistema de IA (niveles de servicio,

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					configuración, monitorización, incidentes, etc.) El tercero se comprometerá a cumplir los requisitos de transparencia y de colaboración en la gestión de seguridad del modelo impuestos por la entidad.
op.ext.3	Protección de la cadena de suministro		X	Compartida	La entidad deberá analizar el impacto que puede tener sobre su sistema un incidente accidental o deliberado que tenga su origen en un proveedor del sistema de IA y esta dependencia se tendrá en cuenta en el plan de continuidad de la entidad. A su vez el tercero cumplirá las obligaciones como proveedor impuestas en el RIA para garantizar la seguridad del sistema de IA.
op.ext.4	Interconexión de sistemas		X	Entidad	
op.nub	**Servicios en la nube**				
op.nub.1	Protección de servicios en la nube	X		Tercero	El tercero deberá ser conforme con el ENS o regirse por las guías

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					CCN-STIC relativas a la protección de servicios de IA en la nube.
op.mon	**Monitorización del sistema**				
op.mon.1	Detección de intrusión	X		Tercero	
op.mon.3	Vigilancia		X	Compartida	La entidad supervisará y coordinará con el tercero la monitorización de las medidas de configuración de la seguridad. El tercero llevará un registro de la configuración de seguridad y los usuarios activos, volúmenes, fuentes consultadas, errores, bloqueos, consumo, disponibilidad, conexiones y eventos anómalos. En los casos establecidos, como por ejemplo, datos de entrada sospechosos, errores en la formulación y evaluación de las pruebas, el tercero notificará a la entidad para que esta revise y solicite la intervención en el sistema de IA en caso necesario.
mp	**Medidas de protección**				
mp.per	**Gestión del personal**				
mp.per.3	Concienciación	X		Entidad	

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
mp.per.4	Formación	X		Entidad	
mp.sw	**Protección de las aplicaciones informáticas**				
mp.sw.1	Desarrollo de aplicaciones		X	Compartida	La entidad trasladará al tercero las pautas de desarrollo seguro vigentes (separación de entornos, desarrollo seguro, datos de pruebas). El tercero comunicará las pautas de desarrollo seguro que se utilicen.
mp.info	**Protección de la información**				
mp.info.1	Datos personales	X		Entidad	La entidad garantizará el cumplimiento art. 28 del RGPD [Ref.2] y, cuando proceda, EIPD, transferencias, subencargados, conservación, etc. Si las evaluaciones pudieran dar lugar a decisiones exclusivamente automatizadas con efectos jurídicos o significativamente similares, habría que contemplar también el art. 22 RGPD.
mp.info.6	Copia de seguridad	X		Compartida	La entidad trasladará los requisitos de copias de seguridad (tiempos y puntos de recuperación objetivo) al tercero e implementará las copias de sus

Código	Medida del ENS	Transversal	Alto riesgo y/o sistema autónomo	Responsable de implantación	Criterio de implantación
					propios datos, asegurando la recuperación en caso necesario. El tercero implementará y comunicará a la entidad el procedimiento de copia de seguridad, asegurando la recuperación en caso necesario.
mp.s	**Protección de los servicios**				
mp.s.2	Protección de servicios y aplicaciones web	X		Compartida	La entidad trasladará los requisitos de seguridad del portal web. El tercero garantizará la seguridad del portal web, teniendo en cuenta las amenazas específicas asociadas a los sistemas de IA (prompt injection, generación de respuestas no autorizadas, evasión de las restricciones de seguridad establecidas para el sistema, etc.).

Tabla 4 Medidas caso de uso sistema autónomo