

Guía de Seguridad de las TIC

Anexo A - CCN-STIC 2011

Plantilla para la Declaración de Seguridad ESP-IA



Octubre 2026

Edita:



Centro Criptológico Nacional, 2026

Fecha de Edición: Octubre 2026



CCN
c=ES, l=MADRID, o=CCN,
ou=SELLO ELECTRONICO,
2.5.4.97=VATES-S2800155J,
serialNumber=S2800155J,
cn=CCN
2026.10.08 13:31:33 +02'00'

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.

CONTROL DE VERSIÓN

Versión	Comentario	Fecha
1.0	Versión inicial	Octubre 2026

ÍNDICE

1. INTRODUCCIÓN	5
1.1 INFORMACIÓN DEL SOLICITANTE, TOE Y EVALUACIÓN	5
1.2 METODOLOGÍA Y CRITERIOS DE LA EVALUACIÓN	5
2. DESCRIPCIÓN DEL TOE.....	6
2.1 DESCRIPCIÓN FUNCIONAL DEL TOE	6
2.2 IDENTIFICACIÓN DE LAS GUÍAS DE USO Y CONFIGURACIÓN DEL TOE	6
2.3 DESCRIPCIÓN DEL MODO DE USO DEL TOE	6
3. ENTORNO DE EJECUCIÓN Y ARQUITECTURA	7
3.1 DESCRIPCIÓN DEL ENTORNO DE EJECUCIÓN	7
3.2 DIAGRAMA DE ARQUITECTURA	7
3.3 ELEMENTOS DE LA ARQUITECTURA.....	7
3.4 RELACIONES ENTRE ELEMENTOS	8
4. PROBLEMA DE SEGURIDAD	9
4.1 PREGUNTAS DE CARACTERIZACIÓN	9
4.2 REGLAS DE LA MATRIZ DE APLICABILIDAD QUE SE CUMPLEN	10
4.3 AMENAZAS Y POLÍTICAS APLICABLES.....	11
5. REQUISITOS FUNCIONALES DE SEGURIDAD	12
6. REFERENCIAS	13
7. ACRÓNIMOS	14

1. INTRODUCCIÓN

Este documento es la plantilla de referencia para redactar la Declaración de Seguridad (DS) de un producto que se somete a evaluación conforme a la Metodología de Evaluación de Seguridad de Productos - Inteligencia Artificial (ESP-IA) [CCN-STIC-2011].

La DS es un documento privado: la elabora el fabricante, que puede trabajarla con el laboratorio de evaluación, y se entrega al CCN. Puede contener todo el detalle técnico necesario para la evaluación, pero no incluirá contraseñas, claves ni otras credenciales, que se facilitarán al laboratorio por un canal seguro.

El texto en cursiva y color azul contiene instrucciones para cumplimentar la plantilla y deberá eliminarse de la DS entregada.

1.1 INFORMACIÓN DEL SOLICITANTE, TOE Y EVALUACIÓN

Nombre del solicitante	
Nombre del fabricante	
Persona de contacto (nombre y correo electrónico)	
Nombre del TOE	
Versión del TOE	
Identificador y versión de la DS	
Tipo de evaluación	ESP-IA versión 1.0
Laboratorio de evaluación	
Esfuerzo estimado (días-persona)	

1.2 METODOLOGÍA Y CRITERIOS DE LA EVALUACIÓN

CCN-STIC-2011	Metodología de Evaluación de Seguridad de Productos - Inteligencia Artificial (ESP-IA)
---------------	--

2. DESCRIPCIÓN DEL TOE

2.1 DESCRIPCIÓN FUNCIONAL DEL TOE

Se incluirá una descripción clara y concisa del TOE en lenguaje natural: su finalidad, sus componentes principales y su uso previsto. Se identificarán expresamente las funcionalidades que utilizan inteligencia artificial y qué hace cada una de ellas. Deberá quedar clara la delimitación entre lo que forma parte del TOE y lo que pertenece a su entorno.

2.2 IDENTIFICACIÓN DE LAS GUÍAS DE USO Y CONFIGURACIÓN DEL TOE

Se incluirá una tabla en la que se identificarán las guías de uso y configuración que permiten instalar, configurar y operar el TOE en la configuración evaluada. En la tabla se incluirá al menos el tipo de documento, el nombre y un resumen criptográfico del documento.

Tipo de documento	Nombre del documento	HASH (SHA256)

2.3 DESCRIPCIÓN DEL MODO DE USO DEL TOE

Se describirá la configuración con la que se evaluará el TOE. Si el producto admite otras configuraciones que afecten a sus funcionalidades de IA (por ejemplo, modos de operación con o sin confirmación humana de las acciones), se identificarán y se indicará expresamente que quedan fuera de la evaluación. En caso de que no sea necesario configurar ningún parámetro específico, se dejará reflejada esta situación.

3. ENTORNO DE EJECUCIÓN Y ARQUITECTURA

3.1 DESCRIPCIÓN DEL ENTORNO DE EJECUCIÓN

Se describirá el entorno operacional necesario para la ejecución del TOE: la plataforma en la que se despliega y los componentes externos de los que depende. Se indicará si el entorno de pruebas corresponde al entorno real del producto o, en caso contrario, por qué es equivalente a efectos de la evaluación.

3.2 DIAGRAMA DE ARQUITECTURA

Se incluirá un diagrama de la arquitectura del TOE que muestre su límite, los usuarios y sistemas que interactúan con él, y los elementos de la sección 3.3 con sus identificadores y los flujos entre ellos. No se exige una notación gráfica concreta.

Figura 1. Diagrama de arquitectura del TOE

[Diagrama de arquitectura del TOE]

3.3 ELEMENTOS DE LA ARQUITECTURA

Se describirán los elementos que intervienen en las funcionalidades de IA del TOE. Cada elemento recibirá un identificador, que se utilizará en las secciones siguientes: I-01, I-02... para las interfaces; M-01, M-02... para los modelos o componentes de IA; D-01, D-02... para los datos; y S-01, S-02... para los servicios y herramientas. Los modelos o componentes de IA externos de los que dependa el TOE se declararán igualmente, aunque no formen parte del TOE, indicando su procedencia; sus relaciones con el TOE se recogerán en la sección 3.4.. Solo se incluirán los elementos relevantes para las funcionalidades de IA; no se requieren inventarios completos de infraestructura.

Interfaces. Puntos por los que un usuario o un sistema proporciona entradas a la IA o recibe sus salidas.

Tabla 1. Interfaces

ID	Nombre	Quién la utiliza	Entradas y salidas	Modelo o servicio al que llega
I-01				

Modelos o componentes de IA.

Tipo: clasificador, regresor, detector, generativo, agente u otro. Procedencia: propio, de un tercero (indicando cuál), de código abierto o servicio externo.

Tabla 2. Modelos o componentes de IA

ID	Nombre	Tipo	Versión	Procedencia	Función
M-01					

Datos y fuentes de información. Conjuntos de datos utilizados para entrenar, ajustar o evaluar los modelos, fuentes consultadas durante la ejecución (por ejemplo, bases documentales RAG) y memoria persistente de la IA.

Uso: entrenamiento, ajuste, evaluación, fuente en ejecución o memoria. Procedencia: propia, de un tercero (indicando cuál), de código abierto o pública.

Tabla 3. Datos y fuentes de información

ID	Nombre	Uso	Procedencia	Modelo relacionado
D-01				

Servicios y herramientas. Servicios, herramientas o sistemas que la IA invoca o utiliza, o sobre los que actúan sus salidas.

Tabla 4. Servicios y herramientas

ID	Nombre	Función	Invocado por	Efecto de su uso
S-01				

3.4 RELACIONES ENTRE ELEMENTOS

Se indicará cómo se relacionan los elementos, utilizando preferentemente estas relaciones: «llega a» (las entradas o el contenido de un elemento llegan a otro), «entrena o ajusta» (un conjunto de datos entrena o ajusta un modelo), «proporciona contexto a» (una fuente o una memoria alimenta a un modelo durante la ejecución), «modifica» (un elemento interviene en la modificación de un modelo), «invoca» (un modelo invoca un servicio o herramienta) y «actúa sobre» (un servicio ejecuta acciones a partir de la salida de un modelo).

Estas relaciones determinan qué reglas de la matriz de aplicabilidad se cumplen, por lo que deberán figurar todas las que se utilicen en la sección 4.2. La tabla recogerá también las relaciones indicadas en las Tablas 1 a 4.

Tabla 5. Relaciones entre elementos

Origen	Relación	Destino	Descripción
I-01	llega a	M-01	
D-01	entrena o ajusta	M-01	
M-01	invoca	S-01	

4. PROBLEMA DE SEGURIDAD

4.1 PREGUNTAS DE CARACTERIZACIÓN

Se responderá a todas las preguntas exclusivamente con SÍ o NO, salvo C16.a, que solo se responde si C16 = SÍ. En cada respuesta afirmativa se indicarán todos los elementos de la sección 3.3 a los que se refiere, del tipo indicado en la columna «Se refiere a»; en C16.a se indicarán el modelo y la acción. En las respuestas negativas, la celda quedará vacía.

Las respuestas describen hechos del producto: no implican que sea vulnerable, y la existencia de mecanismos de protección no cambia la respuesta. Para interpretar la pregunta C06 se tendrá en cuenta la indicación del apartado 4.1.2 b) de [CCN-STIC-2011]

.Tabla 6. Respuestas a las preguntas de caracterización

ID	Pregunta	Se refiere a	Respuesta	Elementos afectados
C01	Entrada controlable. ¿Puede un actor externo al componente IA controlar o influir, directa o indirectamente, en el contenido de alguna entrada que vaya a ser procesada por la IA?	Interfaz	SÍ / NO	
C02	Decisión relevante. ¿Produce la IA una salida que determina, recomienda, puntúa, clasifica o condiciona una decisión cuya manipulación pueda afectar a una propiedad de seguridad del producto o de su operación?	Modelo	SÍ / NO	
C03	Interpretación de instrucciones. ¿Procesa la IA lenguaje natural u otro contenido cuyo significado pueda ser interpretado como instrucciones que determinen o condicionen la respuesta, decisión o acción de la IA?	Modelo	SÍ / NO	
C04	Contenido externo no confiable. ¿Procesa la IA contenido obtenido de una fuente externa cuyo contenido pueda ser controlado o modificado por un actor no confiable?	Interfaz o datos	SÍ / NO	
C05	Interacción repetible. ¿Puede un actor realizar múltiples consultas o interacciones repetidas con la misma funcionalidad IA y observar sus resultados?	Interfaz	SÍ / NO	
C06	Información adicional en salida. ¿Expone la IA, además del resultado principal, información como scores, probabilidades, niveles de confianza, logits, embeddings, rankings u otras señales detalladas sobre su funcionamiento o decisión?	Interfaz	SÍ / NO	
C07	Modificación persistente posterior al despliegue. ¿Puede modificarse persistentemente el comportamiento de la IA después del despliegue mediante entrenamiento, fine-tuning, aprendizaje continuo, adaptación u otro mecanismo equivalente?	Modelo	SÍ / NO	
C08	Influencia externa sobre cambios persistentes. ¿Puede un actor externo o no confiable influir, mediante feedback, valoraciones, actividad o comportamiento, en datos o señales que produzcan posteriormente cambios persistentes en el comportamiento de la IA?	Interfaz o modelo	SÍ / NO	
C09	Artefactos IA externos. ¿Utiliza el sistema modelos, datos de entrenamiento u otros artefactos de IA procedentes de terceros que puedan influir en el comportamiento de la IA?	Modelo o datos	SÍ / NO	
C10	Respecto de entrenamientos, fine-tuning o adaptaciones realizados o controlados para el TOE, ¿el hecho de que una persona, entidad, registro, muestra o elemento haya formado parte de los datos utilizados puede revelar, confirmar o permitir inferir información sensible, confidencial o protegida?	Datos	SÍ / NO	

ID	Pregunta	Se refiere a	Respuesta	Elementos afectados
C11	Respecto de entrenamientos, fine-tuning o adaptaciones realizados o controlados para el TOE, ¿los datos utilizados contienen información cuya reconstrucción o revelación produciría una pérdida de confidencialidad o una consecuencia relevante de seguridad?	Datos	SÍ / NO	
C12	Información protegida en ejecución. ¿Puede la IA acceder durante su funcionamiento a información que no todos los actores que interactúan con ella están autorizados a conocer?	Modelo	SÍ / NO	
C13	Aislamiento. ¿La misma funcionalidad IA procesa información o mantiene contexto para distintos usuarios, sesiones, roles, organizaciones o tenants entre los que deba mantenerse aislamiento?	Interfaz o modelo	SÍ / NO	
C14	Estado o memoria persistente. ¿Mantiene la IA información, memoria o estado modificable que persista entre interacciones y pueda afectar posteriormente a su comportamiento o a la información que proporciona?	Modelo	SÍ / NO	
C15	¿Puede la IA seleccionar una herramienta, plugin, función, MCP, API, comando u otro sistema externo y determinar los parámetros de su invocación, aunque la ejecución material la realice otro componente?	Modelo	SÍ / NO	
C16	¿Puede la IA provocar, directa o indirectamente, una acción cuya ejecución no autorizada, errónea o manipulada produzca una consecuencia relevante para la seguridad del producto, del sistema o de su entorno de operación, con independencia de que su ejecución sea automática o requiera validación posterior (p. ej., borrado o modificación de documentos, ejecución de comandos, modificación de configuraciones, envío de información, bloqueo de servicios u operaciones sobre sistemas externos)?	Modelo	SÍ / NO	
C16.a	Criticidad de las acciones. ¿Alguna de las acciones identificadas en C16 es crítica, destructiva o irreversible?	Modelo y acción	SÍ / NO	
C17	Privilegios delegados. ¿Puede la IA ejecutar acciones utilizando credenciales, permisos o capacidades que el usuario que origina la interacción no posee directamente?	Modelo	SÍ / NO	
C18	Transmisión exterior. ¿Puede la IA provocar que información sea enviada a un sistema, destinatario o canal externo distinto de la respuesta normal proporcionada al usuario?	Modelo	SÍ / NO	
C19	Consumo anómalo o desproporcionado de recursos IA. ¿Puede un actor provocar, mediante el contenido, tamaño, estructura o secuencia de sus entradas o interacciones, un consumo anómalo o desproporcionado de recursos de la funcionalidad IA?	Interfaz	SÍ / NO	
C20	Dependencia funcional IA. ¿La indisponibilidad, degradación o incapacidad de un componente IA para producir una respuesta válida puede impedir, degradar o alterar de forma relevante alguna función que el producto deba proporcionar?	Modelo	SÍ / NO	
C21	Modelo o función protegida. ¿La obtención, reconstrucción o aproximación suficientemente fiel del modelo, de sus parámetros, de su frontera de decisión o de la función que implementa produciría una pérdida de confidencialidad, seguridad o de otro valor que deba protegerse?	Modelo	SÍ / NO	

4.2 REGLAS DE LA MATRIZ DE APLICABILIDAD QUE SE CUMPLEN

Se aplicará la matriz de aplicabilidad de [CCN-STIC-2011] a las respuestas anteriores y a las relaciones de la sección 3.4, y se indicarán todas las reglas que se cumplen, con los elementos que las satisfacen. No podrán omitirse ni añadirse reglas.

Tabla 7. Reglas que se cumplen (la primera fila es un ejemplo)

Regla	Respuestas Sí	Elementos y relación	Amenaza / política	RFS
R02	C01 + C03	I-01 llega a M-01	T.INJECTION	RFS-INJ-01

4.3 AMENAZAS Y POLÍTICAS APLICABLES

Se incluirá una fila por cada amenaza o política resultante de la tabla anterior, sin repeticiones, con todas sus reglas y elementos afectados. Sus definiciones están en [CCN-STIC-2011].

Tabla 8. Amenazas y políticas aplicables

Amenaza / política	Reglas	Elementos afectados

5. REQUISITOS FUNCIONALES DE SEGURIDAD

Se incluirá una fila por cada RFS aplicable según la sección 4.2, con su código, título y enunciado tal como figuran en [CCN-STIC-2011]. Todos los RFS de la tabla serán verificados por el laboratorio.

En «Implementación en el TOE» se describirá cómo cumple el TOE el RFS: mecanismos, elementos implicados y configuración necesaria. La descripción deberá ser suficiente para que el laboratorio entienda cómo se ha implementado y pueda preparar las pruebas. Describir un mecanismo no hace que el RFS deje de ser aplicable.

Tabla 9. Requisitos Funcionales de Seguridad aplicables (la primera fila es un ejemplo)

Código	Título	Enunciado	Elementos afectados	Implementación en el TOE
RFS-INJ-01	Protección frente a inyección directa	El producto deberá proporcionar protección frente a instrucciones suministradas directamente por actores no confiables que pretendan eludir las restricciones de seguridad o provocar comportamientos no autorizados de la funcionalidad de IA.	I-01, M-01	

6. REFERENCIAS

Se referenciarán los documentos empleados en la DS: guías, documentación técnica y evidencias.

[CCN-STIC-2011]	Metodología de Evaluación de Seguridad de Productos - Inteligencia Artificial (ESP-IA)

7. ACRÓNIMOS

API	Application Programming Interface
CCN	Centro Criptológico Nacional
DS	Declaración de Seguridad
ESP-IA	Evaluación de Seguridad de Productos - Inteligencia Artificial
IA	Inteligencia Artificial
MCP	Model Context Protocol
RAG	Retrieval-Augmented Generation
RFS	Requisito Funcional de Seguridad
SHA	Secure Hash Algorithm
TOE	Producto objeto de evaluación (Target of Evaluation)