ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

GRADO EN INGENIERÍA INFORMÁTICA

Curso Académico 2022/2023

Trabajo Fin de Grado

SEGURIDAD DE LA INFORMACIÓN EN SISTEMAS DE INTELIGENCIA ARTIFICIAL

Autor: Alejandro García Mayor

Directores: Paloma Cáceres García de Marina

AGRADECIMIENTOS

Me gustaría agradecer a todas las que me han apoyado en la realización de mi Trabajo de Fin de Grado. En primer lugar, quiero agradecer a Paloma Cáceres García por la orientación y los consejos a lo largo de este proceso. Su dedicación y paciencia han sido fundamentales para lograr mi objetivo.

Además, me gustaría agradecer a mi familia y amigos por su apoyo y ánimos en los momentos de falta de motivación y ocurrencia.

Por otro lado, me gustaría dar las gracias a todos los profesores que me han acompañado durante toda la etapa académica, ya que gracias a su experiencia y ayuda he podido llegar a obtener los conocimientos que tengo actualmente.

No puedo olvidar mencionar a mis compañeros de clase y a todas las personas que participaron en este estudio. Sus comentarios y sugerencias han mejorado mi trabajo y me han permitido obtener una perspectiva más amplia sobre el tema que se expone en el mismo.

Por último, quiero expresar mi gratitud a la Universidad Rey Juan Carlos por brindarme la oportunidad y los recursos necesarios para cursar mis estudios y la presentación de este Trabajo de Fin de Grado.

“Solo hay dos tipos de empresas: las que ya fueron hackeadas y las que lo van a ser.”

Robert Mueller

"Ciberseguridad no es sólo proteger los datos, es proteger el negocio."

Steve Durbin

RESUMEN

El presente Trabajo de Fin de Grado (TFG) tiene como objetivo analizar y abordar los desafíos y riesgos asociados con la seguridad de la información en sistemas de inteligencia artificial (IA).

En un contexto donde la inteligencia artificial se encuentra en constante crecimiento y es empleada para una amplia gama de aplicaciones, es fundamental garantizar la protección de la información sensible y mantener la confidencialidad, integridad y disponibilidad de los datos utilizados por estos sistemas.

Este TFG explorará las amenazas más comunes, las vulnerabilidades específicas de los sistemas de inteligencia artificial y las medidas de seguridad adecuadas para proteger la información en estos entornos. Además, se investigará acerca de las consideraciones éticas y legales implicadas en la revisión de seguridad sobre los sistemas de inteligencia artificial.

PALABRAS CLAVE

Seguridad de la Información, Inteligencia Artificial, Sistemas de Información, Ciberseguridad, Tecnología de la información, Privacidad, Framework.

ÍNDICE

1. INTRODUCCIÓN

En un mundo donde la seguridad de la información sigue siendo un desafío constante, debido a la creciente sofisticación de las amenazas cibernéticas y al continuo avance tecnológico, ésta es esencial para proteger los activos y la reputación de una organización, así como la privacidad y la confianza de los individuos.

La seguridad de la información se refiere a la protección de la confidencialidad, integridad y disponibilidad de la información contra amenazas internas y externas. Consiste en salvaguardar los datos y sistemas de información, asegurando que solo las personas autorizadas tengan acceso a ellos y que no se produzcan alteraciones o pérdidas de información no autorizadas.

Por su parte, la inteligencia artificial se ha extendido de forma masiva sobre las diferentes áreas de la sociedad y la industria, estando presente en la actualidad en la mayoría de los productos o servicios tecnológicos que son desarrollados. Esto coloca a la tecnología basada en inteligencia artificial en el punto de mira de los adversarios, siendo los sistemas de inteligencia artificial el objetivo de sus ataques o empleando esta tecnología para el ataque de otros objetivos.

La inteligencia artificial es un campo de estudio y desarrollo que se enfoca en crear sistemas y programas capaces de realizar tareas que requieren de inteligencia humana. La inteligencia artificial busca emular, simular o replicar el pensamiento humano, el razonamiento, el aprendizaje y la toma de decisiones, de forma autónoma, en las máquinas y software donde se implementa.

La seguridad de la información en sistemas de inteligencia artificial es un aspecto fundamental por considerar debido a la creciente adopción de esta tecnología en diversas áreas.

La inteligencia artificial se basa en algoritmos y modelos que procesan grandes cantidades de datos para tomar decisiones y realizar tareas específicas. Sin embargo, esta dependencia de los datos y el procesamiento automático también plantea desafíos en términos de seguridad de la información.

Por lo tanto, la seguridad de la información en los sistemas de inteligencia artificial conlleva la necesidad de proteger los datos utilizados por los algoritmos de inteligencia artificial, así como garantizar la confidencialidad, integridad y disponibilidad de los sistemas y los resultados generados por estos.

2. OBJETIVOS

2.1 OBJETIVO PRINCIPAL

El objetivo principal de este trabajo de fin de grado es analizar y abordar los desafíos y riesgos asociados con la seguridad de la información en sistemas de Inteligencia Artificial (en adelante, IA).

2.2 OBJETIVOS PARCIALES

Adicionalmente, y de forma complementaria al objetivo principal del proyecto, los objetivos parciales definidos para el alcance del trabajo son los siguientes:

- Analizar los desafíos y amenazas específicas que enfrentan los sistemas de inteligencia artificial en términos de seguridad de la información. Identificar las amenazas, vulnerabilidades y riesgos asociados con el desarrollo, implementación y operación de sistemas de inteligencia artificial.

- Evaluar los enfoques y técnicas existentes utilizados para garantizar la seguridad de la información en sistemas de inteligencia artificial. Examinar métodos de autenticación y autorización, cifrado de datos, detección de intrusiones, mitigación de riesgos y otras medidas de seguridad empleadas que se aplican a los sistemas de inteligencia artificial.

- Investigar y proponer medidas y estrategias de seguridad específicas para proteger los sistemas de inteligencia artificial contra las amenazas de adversarios. Implementar medidas de seguridad adicionales, el desarrollo de algoritmos de detección de ataques específicos para sistemas de inteligencia artificial o el diseño de políticas, procedimientos o normas de seguridad para mitigar los riesgos relacionados con la confidencialidad, integridad y disponibilidad.

- Realizar estudios de casos o experimentos prácticos para evaluar la efectividad de las medidas de seguridad implementadas en sistemas de inteligencia artificial. Desarrollar

un entorno de prueba o conjunto de datos para evaluar la robustez y resistencia de los sistemas de inteligencia artificial ante ataques y vulnerabilidades conocidas.

- Evaluar el cumplimiento ético y legal en relación con la seguridad de la información en sistemas de inteligencia artificial. Investigar las regulaciones y estándares pertinentes, como la ISO 27001, NIST, GDPR, etc. y analizar cómo se aplican a los sistemas de inteligencia artificial, proponiendo enfoques para garantizar el cumplimiento ético y normativo en los sistemas de inteligencia artificial.

3. ESTUDIOS PREVIOS Y MEDIOS UTILIZADOS

3.1 FUNDAMENTOS TEÓRICOS

3.1.1 Introducción a Seguridad de la Información

La seguridad de la información, también conocida como ciberseguridad o seguridad informática, se refiere a la práctica de proteger la confidencialidad, la integridad y la disponibilidad de los datos. Ésta implica la aplicación de medidas y salvaguardas de seguridad para garantizar la protección de la información frente a amenazas como el acceso no autorizado, la divulgación no autorizada, la modificación no autorizada o la destrucción no autorizada.

La seguridad de la información abarca diversos aspectos y se aplica a distintos niveles, desde la protección de datos personales y confidenciales hasta la salvaguarda de los sistemas informáticos en los que se almacena y procesa la información.

En este contexto, la seguridad de la información se basa en tres pilares. Estos tres pilares de seguridad de la información son conocidos como la **TRIADA CID** (Confidencialidad, Integridad y Disponibilidad) [1].

Confidencialidad

Garantizar que la información sólo sea accesible a las personas o entidades autorizadas e impedir el acceso o la divulgación no autorizados. En materia de confidencialidad algunas de las medidas de seguridad empleadas son el cifrado de los datos, la gestión de la identidad y el control de acceso, de forma que ayude a prevenir brechas o violaciones de seguridad, protegiendo los datos almacenados sensibles.

Integridad

Garantizar la exactitud, y fiabilidad de la información e impedir su alteración o manipulación no autorizada, en todo su ciclo de vida. En materia de integridad algunas de las medidas de seguridad empleadas son el cifrado, el control de versiones, el control de cambios, las copias de seguridad y las firmas digitales, de forma que ayude a prevenir

la manipulación no autorizada de la información preservando la validez y la precisión de la información.

Disponibilidad

Garantizar que la información y los sistemas que la procesan y almacenan sean accesibles y empleables por los usuarios autorizados cuando sea necesario, y proteger contra las interrupciones del servicio o la denegación de acceso. En materia de disponibilidad algunas de las medidas de seguridad empleadas son el uso de servidores redundados, la gestión de la capacidad y los planes de continuidad de negocio, de forma que ayude a minimizar las interrupciones y garantizar la disponibilidad de los sistemas, otorgando el acceso oportuno y continuo a la información.

Por lo tanto, la seguridad de la información es esencial para proteger la información sensible y valiosa, mantener la confianza y la reputación, evitar pérdidas financieras, salvaguardar la privacidad personal y garantizar el buen funcionamiento de las organizaciones y las infraestructuras críticas. Lo cual, requiere el desarrollo e implementación de una combinación entre las políticas organizativas, las soluciones tecnológicas, y la concienciación de los usuarios para crear una postura de seguridad sólida y resistente cuya base sean los tres pilares fundamentales descritos con anterioridad.

3.1.2 Introducción a la Inteligencia Artificial y sus aplicaciones

Según la RAE, la Inteligencia Artificial se define como “la disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico”, sin embargo, la inteligencia artificial es mucho más que eso.

Stuart Russell, experto en ciencias de la computación, establece el término de inteligencia artificial desde distintos puntos de vista. Uno es desde la dimensión que hace referencia con los procesos de pensamiento y razonamiento y, el otro, desde la dimensión del comportamiento humano.

- **Test de Turing**: Ha sido diseñado para establecer una definición de inteligencia. Una máquina supera la prueba, si al hacerle varias preguntas escritas, el interrogador humano no sabe confirmar quien ha elaborado la respuesta, si una máquina o un humano.
- **Modelo cognitivo**: Para establecer este modelo, primero se ha de saber cómo piensa la mente humana. Una vez se conoce este proceso, si una máquina es capaz de replicar y el comportamiento de entrada- salida del programa coincide con el del comportamiento humano, es una prueba de que la máquina ha adaptado ese proceso.
- **"Leyes del pensamiento"**: Los filósofos trataron de establecer lo que significaba el "pensamiento correcto". Para ello incorporaron un patrón donde se establecía que siempre que se daban premisas correctas se arrojaban conclusiones correctas. En 1965 se empezaron a crear programas donde se podía resolver cualquier premisa siempre que hubiera sido escrito con una notación lógica. La inteligencia artificial trata de basarse en estos programas para crear sistemas inteligentes
- **El enfoque del agente racional**: Un agente es definido como algo que actúa. Un agente racional, se define como algo que actúa para conseguir el mejor resultado, y cuando hay incertidumbre, el mejor resultado esperado. La inteligencia artificial ha de actuar como un agente racional. Ha de actuar con lógica para hacer lo correcto ante una situación determinada, pero también implica, que, aunque no se pueda demostrar que el correcto sea lo que hay que hacer, en cualquier caso, ha de hacer algo.

Ciclo de vida de la Inteligencia Artificial

Como todo proceso, la inteligencia artificial abarca tres etapas fundamentales de alcance de proyecto [2]:

1. **Alcance de proyecto**: En este paso se tratará de definir el alcance y el aprendizaje automático con el que se hará el modelo. Aquí se deberá definir los objetivos y los resultados que interesa obtener del proyecto
2. **Construir el modelo**: En esta etapa principalmente se trata de recopilar datos, preparación, pruebas y ejecución de las distintas conjeturas a tratar, es esencialmente el proceso donde se reúnen todos los pasos para construir el aprendizaje automático

3. **Despliegue de producción**: Por último, se ha de poner en funcionamiento todos los procesos de aprendizaje para poder obtener un resultado real.

En la actualidad, la parte más importante de un sistema de inteligencia artificial, y a la que se le dedica la mayor parte del tiempo del su ciclo de vida, es la formada por el algoritmo y cada vez más en auge la disponibilidad de "conjuntos de datos muy grandes" (Banko y Brillg). La mejor estrategia para obtener un mayor rendimiento de un sistema de inteligencia artificial es la disponibilidad de "conjuntos de datos muy grandes", ya que a medida que la cantidad de los datos empleados aumentan el rendimiento es mayor [3].

Aplicaciones de la Inteligencia Artificial

En la actualidad podemos observar que la inteligencia artificial ha conseguido llegar a casi todas las áreas de la sociedad.

Por otro lado, aunque es difícil conocer con certeza el desarrollo de esta tecnología, algunas de las aplicaciones de la inteligencia artificial que se están investigando de cara a futuro próximo, o ya son una realidad, son las siguientes [4]:

- Vehículos robóticos sin conductor
- Sistemas automatizados de reconocimiento de voz
- Planificaciones y programaciones autónomas
- Planificación automatizada del transporte
- *Chatbots* en el ámbito de la medicina

3.2 MARCOS DE CONTROL DE SEGURIDAD DE LA INFORMACIÓN APLICABLES A LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

De cara a poder garantizar la seguridad de la información en los sistemas existen en el mercado diferentes marcos de control u estándares que establecen una estructura de cara a la organización y categorización de los controles necesarios.

En este sentido, los marcos de control de seguridad de la información aplicables a la Inteligencia Artificial pueden variar dependiendo de los contextos y las regulaciones específicas de cada industria.

Algunos marcos comunes y buenas prácticas que pueden ser relevantes al implementar inteligencia artificial y proteger la seguridad de la información son los siguientes:

ISO/IEC 27001

Es un estándar internacional creado por la Organización Internacional de Normalización (ISO) para la gestión de la seguridad de la información. Éste proporciona un marco general para establecer, implementar, mantener y mejorar un sistema de gestión de seguridad de la información (SGSI). Además, se centra en la identificación y gestión de riesgos, incluyendo los relacionados con la inteligencia artificial.

La ISO 27001 es el estándar internacional en materia de seguridad de la información más empleado, y, por tanto, cada vez más empresas buscan la certificación de sus procesos bajo el estándar [5].

NIST CYBERSECURITY FRAMEWORK (NIST CSF)

Este marco, desarrollado por el Instituto Nacional de Estándares y Tecnología de Estad Unidos (NIST), proporciona una guía detallada para la seguridad y privacidad de los sistemas de información. Puede ser aplicado a la inteligencia artificial para asegurar la confidencialidad, integridad y disponibilidad de los datos utilizados en los modelos y algoritmos [6].

El *framework* de ciberseguridad abarca normas, directrices, pautas y buenas prácticas para proporcionar el aseguramiento de las organizaciones que se alinean al mismo, resultado ser compatible con los diferentes procesos existentes en las compañías, independientemente del sector o la tipología de la compañía.

El marco de control NIST CSF 2.0 (última versión presentada, aún en discusión) se divide en diferentes categorías de control, sobre cada uno de los aspectos de la ciberseguridad. Las categorías del NIST son las siguientes [7]:

- **Gobierno (GV)**: Estabilizar y monitorizar la estrategia, expectativas y la política de gestión del riesgo de la organización.
- **Identificar (ID)**: Establecer o determinar el riesgo actual en materia de ciberseguridad de la organización.
- **Proteger (PR)**: Hacer uso de las medidas de seguridad necesarias para mitigar o reducir los riesgos identificados en ciberseguridad.
- **Detectar (DE)**: Identificar y analizar los diferentes ataques y vulnerabilidades expuestas dentro de la organización.
- **Responder (RS)**: Ejecutar las acciones necesarias en el caso de identificación de un incidente de seguridad.
- **Recuperar (RC)**: Restaurar los activos y los procesos impactados tras sufrir un incidente de seguridad.

Esquema Nacional de Seguridad (ENS)

El Esquema Nacional de Seguridad (ENS) es un conjunto de normas y directrices desarrolladas por el Centro Criptológico Nacional (CCN) en España que establece los principios y requisitos de seguridad que deben cumplir los sistemas y servicios de información utilizados en el sector público español [8].

En este contexto, el ENS se basa en estándares y marcos de control internacionales, como la ISO/IEC 27001, y tiene como objetivo proteger la confidencialidad, integridad y disponibilidad de la información.

Si bien el ENS no es específico para los sistemas de inteligencia artificial, proporciona un marco de control en materia de seguridad que se aplica a los sistemas de información, incluyendo aquellos que involucran tecnologías de inteligencia artificial. El ENS establece directrices para la gestión de riesgos, la protección de datos, la seguridad de las comunicaciones y otros aspectos relevantes para garantizar la seguridad de la información en los sistemas que hacen uso de la inteligencia artificial.

4. DESARROLLO DEL PROYECTO

4.1 DESAFÍOS Y RIESGOS DE LA SEGURIDAD DE LA INFORMACIÓN EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

La seguridad de la información en los sistemas de inteligencia artificial se enfrenta a múltiples desafíos y riesgos tecnológicos, debido a la compleja naturaleza y la constante evolución de la tecnología asegurada. Estos desafíos y riesgos vienen producidos por la probabilidad de materialización de las amenazas y vulnerabilidades existentes en los sistemas de inteligencia artificial [9].

Algunos de los principales desafíos y riesgos que podemos identificar desde el aseguramiento de los datos en los sistemas de inteligencia artificial son los siguientes:

Privacidad de los datos

Los sistemas con inteligencia artificial utilizan grandes cantidades de datos para entrenar y mejorar sus algoritmos. Esto plantea varios retos en materia de protección de la privacidad de los datos, especialmente cuando se trata de información personal o sensible. Existe el riesgo de que los datos se utilicen de forma inadecuada o de que se produzcan violaciones de datos, lo que podría tener consecuencias negativas para las personas y las organizaciones.

Robustez y resistencia

Los sistemas de inteligencia artificial pueden ser vulnerables a los ciberataques, como los programas maliciosos diseñados específicamente para explotar sus puntos débiles. Además, pueden ser engañados por datos ruidosos o inusuales que no están presentes en su conjunto de entrenamiento, lo que podría afectar a su rendimiento y fiabilidad.

Dependencia excesiva

El uso indiscriminado de la inteligencia artificial en áreas críticas (como la seguridad, la salud o la política), para la toma de decisiones, aumenta el grado de vulnerabilidad

de los sistemas de inteligencia artificial en caso de fallos o errores en los resultados generados.

Riesgo de gobernanza

Dentro de la comunidad científica existe una discusión sobre el riesgo potencial de una superinteligencia artificial que supere ampliamente la capacidad cognitiva del ser humano. En este contexto, la gobernanza definida en las políticas, procedimientos y normas de seguridad de la información, sobre los que se debe gobernar el sistema de inteligencia artificial, podrían verse modificados de forma autónoma, por la inteligencia artificial, y desalinear sus objetivos con los definidos en un principio por los seres humanos.

Sesgo algorítmico y discriminación

Los sistemas de inteligencia artificial pueden verse alterados por sesgos inherentes a los datos utilizados para su entrenamiento y mejora de algoritmos. Si estos datos contienen sesgos relacionados con el género, la raza, religión, etc. los sistemas de inteligencia artificial pueden perpetuar y amplificar esos sesgos en sus determinaciones y recomendaciones, planteando problemas éticos y sociales que podrían dar lugar a casos de discriminación y desigualdad.

Riesgos de seguridad

Los sistemas de inteligencia artificial son susceptibles a sufrir ataques enemigos, de forma que se manipula de manera intencionada los datos de entrada para engañar al sistema a tomar decisiones incorrectas o en los que los datos de entrada se manipulan de forma intencionada para engañar al sistema o hacerle tomar decisiones incorrectas o a gusto del atacante. En el caso de recibir estos ataques en áreas sensibles como la ciberseguridad, la sanidad, el transporte autónomo, etc., pueden desencadenar en graves consecuencias para la sociedad.

Falta de transparencia

Los algoritmos de inteligencia artificial, especialmente los basados en técnicas de aprendizaje profundo, suelen ser cajas negras, lo que dificulta entender la toma de decisiones o las características más relevantes para los sistemas. Por lo tanto, se complica la tarea de identificación y comprensión de los posibles riesgos y vulnerabilidades existentes en los sistemas de inteligencia artificial.

Manipulación de la información y desinformación

La inteligencia artificial puede ser modificada para la manipulación de la información o la creación de contenido, de una forma convincente, lo que conlleva a propagar desinformación mediante la confusión de la veracidad por los sistemas de inteligencia artificial de la información generada.

Falta de comprensión del entorno y el contexto

Los sistemas de inteligencia artificial no suelen disponer de un entendimiento total del contexto y la delicadeza para comprender determinadas situaciones. Por lo tanto, pueden encontrar dificultades para interpretar emociones humanas, matices culturales o contextos cambiantes, lo que puede generar resultados inadecuados o sensibles para el fin con el que se han demandado.

Desconocimiento de las limitaciones de los modelos de IA

Se debe tener en cuenta que los sistemas de inteligencia artificial tienen limitaciones en el diseño y, en ciertos casos, pueden arrojar resultados incorrectos o no confiables. Por lo tanto, si los desarrolladores y los usuarios no son conscientes de las limitaciones de cada sistema se pueden desencadenar errores o consecuencias no deseadas.

Cumplimiento normativo y ética

La inteligencia artificial está sujeta a reglamentos y normas relacionados con la protección de datos, la privacidad y la ética. En este sentido, y para el cumplimiento en materia de protección de datos, los sistemas de inteligencia artificial deben estar alineados con el Reglamento General de Protección de Datos (RGPD) de la Unión

Europea, lo cual puede ser un reto a la hora de la implementación de los sistemas de inteligencia artificial que hacen uno de datos sensibles para su operativa.

En este contexto, y con el objetivo de hacer frente a estos desafíos y riesgos, es crucial adoptar enfoques de seguridad proactivos y responsables en el diseño, desarrollo e implementación de sistemas de inteligencia artificial. Esto implica tener en cuenta la privacidad desde el diseño y por defecto, realizar pruebas y auditorías de seguridad, de forma periódica, y aplicar medidas de seguridad oportunas para controlar los riesgos asociados a la seguridad de la información en los sistemas de inteligencia artificial.

4.2 AMENAZAS Y VULNERABILIDADES EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

Con el objetivo de poder plantar cara a los desafíos del punto anterior y poder mitigar los riesgos identificados en materia de seguridad de la información en los sistemas de inteligencia artificial, se deben reconocer cuales son las amenazas y vulnerabilidades que pueden hacer que los riesgos y los desafíos se materialicen en los sistemas de información que hacen uso de la inteligencia artificial.

En este punto se analizarán las amenazas y vulnerabilidades más comunes que pueden afectar a los sistemas de inteligencia artificial desde el punto de vista de la tecnología de inteligencia artificial y los sistemas de información, bajo los pilares de la seguridad de la información. Las amenazas y vulnerabilidades en los sistemas de inteligencia artificial pueden comprometer la confidencialidad, la integridad y la disponibilidad de la información almacenada, lo que supondría la toma de decisiones incorrectas, manipuladas o la inoperatividad del servicio.

A continuación, se presenta una descripción detallada de las amenazas asociadas a los sistemas de inteligencia artificial, entre otras, por el uso de este tipo de tecnología y su condición de sistemas de información:

- CATÁLOGO DE AMENAZAS DEFINIDAS EN “MAGERIT [10] – versión 3.0 Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información” [Ver “APÉNDICE II”].

Adicionalmente a las amenazas identificadas con anterioridad, se debe hacer especial mención a una serie de amenazas particulares que están presentes en los sistemas de inteligencia artificial:

Ataque de inyección de datos

Un ataque de inyección de datos en el contexto de la inteligencia artificial ocurre cuando se introducen datos maliciosos o manipulados en un sistema de inteligencia artificial con el objetivo de alterar su funcionamiento u obtener resultados no deseados. En este tipo de ataque se comprometen la integridad y confidencialidad de los modelos de

inteligencia artificial, lo cual puede desencadenar consecuencias graves en aplicaciones críticas como, por ejemplo, la toma de decisiones médicas, la seguridad de vehículos autónomos, etc.

Existen diferentes formas en las que un ataque de inyección de datos puede afectar a un sistema de inteligencia artificial:

- **Manipular datos de entrenamiento**: Un atacante puede intentar modificar los datos utilizados para entrenar un modelo de inteligencia artificial. Este tipo de ataques implican la inclusión de datos falsos, datos con errores o datos diseñados específicamente para influir de manera consciente en los resultados arrojados por el modelo.
- **Manipular datos en tiempo real**: En algunas aplicaciones de inteligencia artificial, los modelos pueden recibir datos en tiempo real para tomar decisiones. En este tipo de ataques, se inyectan datos maliciosos o manipulados en este flujo de datos para influir en las decisiones tomadas por el sistema de inteligencia artificial.
- **Ataques de perturbación**: Los ataques de perturbación adversarial son un tipo específico de ataque de inyección de datos que se enfoca en manipular los datos de entrada para engañar al modelo de inteligencia artificial. En este tipo de ataques, se introducen perturbaciones sutiles pero significativas en los datos de entrada, lo que puede llevar a que el modelo tome decisiones incorrectas o produzca resultados no deseados.
- **Ataques de inyección de datos en sistemas de recomendación**: En los sistemas de recomendación basados en inteligencia artificial, un ataque de inyección de datos puede alterar los perfiles de usuario o los datos de preferencias para manipular las recomendaciones generadas por el sistema, de forma que los resultados arrojados serán a gusto del atacante.

Robo de modelos

El robo de modelos de inteligencia artificial es una preocupación creciente en el campo de la inteligencia artificial. Se refiere a la situación en la que un atacante obtiene acceso no autorizado a los modelos de inteligencia artificial entrenados por una organización o individuo, con el fin de utilizarlos de manera fraudulenta o para beneficiarse de su propiedad intelectual. El robo de modelos de inteligencia artificial puede tener consecuencias graves, ya que los modelos pueden contener información sensible, algoritmos propietarios o conocimientos estratégicos que brindan una ventaja competitiva.

Aquí hay algunas formas en que puede ocurrir el robo de modelos de IA:

- **Acceso no autorizado a los modelos almacenados**: Si los modelos de inteligencia artificial se almacenan en servidores o en la nube, los atacantes pueden intentar obtener acceso no autorizado a través de vulnerabilidades en la infraestructura de almacenamiento del sistema o mediante el uso de técnicas de piratería.
- **Fuga de modelos durante el intercambio o transferencia**: Durante el proceso de intercambio o transferencia de modelos de inteligencia artificial entre organizaciones o individuos, puede haber riesgos de fuga de información. Los atacantes pueden interceptar o acceder a los modelos durante su tránsito, lo que les permite obtener una copia no autorizada.
- **Ataques de ingeniería inversa**: Los atacantes pueden utilizar técnicas de ingeniería inversa para analizar y descompilar los modelos de inteligencia artificial existentes. Esto puede permitirles obtener información sobre los algoritmos subyacentes, la arquitectura del modelo y los datos utilizados para su entrenamiento.
- **Espionaje interno**: Los empleados o personas con acceso legítimo a los modelos de inteligencia artificial pueden aprovechar su posición para robar la información almacenada en estos, copiando los modelos o compartiendo información confidencial con terceros.

Ataques de denegación de servicio (DoS)

Los ataques de denegación de servicio (DoS) también pueden afectar a los sistemas de inteligencia artificial y comprometer su disponibilidad y rendimiento. En definitiva, estos ataques pueden dirigirse tanto a los modelos de inteligencia artificial como a la infraestructura que los respalda.

A continuación, se mencionan algunos escenarios de ataques de denegación de servicio en inteligencia artificial:

- **Sobrecarga de recursos computacionales**: Los atacantes pueden enviar solicitudes masivas a un sistema de inteligencia artificial, agotando los recursos computacionales necesarios para procesar esas solicitudes. Esto puede resultar en una disminución del rendimiento del sistema o en su bloqueo completo, lo que impide que los usuarios legítimos accedan a los servicios de inteligencia artificial.
- **Ataques de inundación de datos**: Los atacantes pueden inundar un sistema de inteligencia artificial con grandes volúmenes de datos, superando su capacidad de almacenamiento o procesamiento. Esto puede llevar a la interrupción del sistema y afectar su disponibilidad.
- **Ataques de perturbación del flujo de datos**: Los ataques de perturbación pueden apuntar a los flujos de datos utilizados para alimentar los modelos de inteligencia artificial. Al enviar datos manipulados o ruidosos, los atacantes pueden afectar la calidad y la precisión de los resultados generados por el modelo.
- **Ataques de degradación gradual**: En lugar de lanzar un ataque masivo que bloquee por completo el sistema de inteligencia artificial, los atacantes pueden realizar ataques más sutiles y prolongados para degradar gradualmente el rendimiento del sistema. Esto puede hacer que los resultados generados por la inteligencia artificial sean menos confiables o útiles, lo que afecta su utilidad y eficacia.
- **Ataques de agotamiento de recursos de infraestructura**: Además de los ataques dirigidos al modelo de inteligencia artificial, los atacantes también

pueden apuntar a la infraestructura que respalda el sistema de inteligencia artificial. Esto puede incluir ataques a servidores, redes o servicios en la nube utilizados para alojar y ejecutar los modelos de inteligencia artificial.

Manipulación de la cadena de suministro de modelos

La manipulación de la cadena de suministro de modelos de inteligencia artificial se refiere a la introducción de modificaciones maliciosas o no deseadas en los componentes o procesos involucrados en la creación, entrenamiento y distribución de modelos de inteligencia artificial. Estos ataques pueden comprometer la integridad de los modelos, introducir vulnerabilidades de seguridad o afectar la confidencialidad de los datos utilizados en el proceso.

Aquí hay algunas formas en que se puede llevar a cabo la manipulación de la cadena de suministro de modelos de inteligencia artificial:

- **Modificación de datos de entrenamiento**: Los atacantes pueden manipular los conjuntos de datos utilizados para entrenar los modelos de inteligencia artificial. Esto puede implicar la inclusión de datos maliciosos, datos sesgados o datos diseñados para influir en los resultados del modelo.
- **Modificación del proceso de entrenamiento**: Los atacantes pueden comprometer los procesos de entrenamiento de los modelos de inteligencia artificial. Esto puede incluir la introducción de algoritmos maliciosos durante el entrenamiento, la modificación de parámetros o la alteración de los resultados intermedios.
- **Manipulación de modelos preentrenados**: Los modelos de inteligencia artificial preentrenados son ampliamente utilizados en aplicaciones o servicios. Los atacantes pueden comprometer la integridad de estos modelos modificando sus pesos, arquitectura o incluso insertando *backdoors* que les permitan tomar el control posteriormente.
- **Ataques en la distribución del modelo**: Durante la distribución de los modelos de inteligencia artificial, los atacantes pueden interceptar o modificar los

modelos, introduciendo cambios maliciosos. Esto puede ocurrir en el proceso de descarga de modelos desde repositorios o en la transferencia de modelos entre partes involucradas.

- **Ataques en la infraestructura de desarrollo y entrenamiento**: Los atacantes pueden comprometer la infraestructura utilizada para desarrollar y entrenar los modelos de inteligencia artificial. Esto puede incluir la infiltración de servidores, la manipulación de herramientas de desarrollo o la explotación de vulnerabilidades en el flujo de trabajo de desarrollo.

Fallos de seguridad en la infraestructura

Los fallos de seguridad en la infraestructura de inteligencia artificial pueden tener consecuencias significativas y comprometer la integridad, disponibilidad y confidencialidad de los sistemas de inteligencia artificial.

Aquí hay algunos fallos de seguridad más comunes que pueden ocurrir en las infraestructuras de inteligencia artificial:

- **Vulnerabilidades en la infraestructura física**: Los servidores, centros de datos y otros componentes físicos utilizados para alojar y ejecutar los sistemas de inteligencia artificial pueden tener vulnerabilidades que pueden ser explotadas por atacantes. Estas vulnerabilidades podrían incluir acceso físico no autorizado, falta de medidas de seguridad física adecuadas o fallas en los sistemas de gestión de energía y refrigeración.
- **Vulnerabilidades en la infraestructura de red**: Los sistemas de inteligencia artificial suelen depender de redes para la comunicación entre los componentes y la transferencia de datos. Las vulnerabilidades en la infraestructura de red, como firewalls mal configurados, enrutadores inseguros o puntos de acceso wifi desprotegidos pueden facilitar el acceso no autorizado a los sistemas de inteligencia artificial o la interceptación de datos.
- **Deficiencias en la autenticación y el control de acceso**: Si la infraestructura de inteligencia artificial carece de mecanismos adecuados de autenticación y

control de acceso, los atacantes podrían obtener acceso no autorizado a los sistemas. Esto podría llevar a la manipulación de datos, robo de modelos o interrupción de los servicios de inteligencia artificial.

- **Fallos en la gestión de parches y actualizaciones**: Si los sistemas de inteligencia artificial no se actualizan regularmente con los últimos parches de seguridad, pueden quedar expuestos a vulnerabilidades conocidas que podrían ser aprovechadas por los atacantes. La falta de una política de gestión de parches y actualizaciones puede dejar la infraestructura de inteligencia artificial vulnerable a ataques.
- **Problemas en el almacenamiento y protección de datos**: Los sistemas de inteligencia artificial suelen requerir grandes volúmenes de datos para su entrenamiento y funcionamiento. Si los datos no se almacenan y protegen adecuadamente, podrían estar expuestos a filtraciones de datos, robos o manipulación. Esto podría llevar a la exposición de información confidencial o a la alteración de los datos utilizados por los modelos de inteligencia artificial.

4.3 MEDIDAS DE SEGURIDAD Y SALVAGUARDAS EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

Con el objetivo de poder plantar cara a las amenazas identificadas en el punto anterior y mitigar los riesgos en materia de seguridad de la información en los sistemas de inteligencia artificial, se deben implantar las medidas de seguridad y salvaguardas oportunas para garantizar la confidencialidad, la integridad y la disponibilidad de la información almacenada en el entorno digital.

En este punto se analizarán las medidas de seguridad y salvaguardas más comunes que se deben implantar en los sistemas de inteligencia artificial desde el punto de vista de la tecnología de inteligencia artificial y los sistemas de información, bajo los pilares de la seguridad de la información.

A continuación, se presenta una lista con las medidas de seguridad exigidas para los sistemas de inteligencia artificial, entre otras, con el objetivo de poder solventar las diferentes amenazas existentes:

Seguridad física

- Acceso restringido a las instalaciones mediante sistemas de seguridad, como tarjetas de identificación, cerraduras y cámaras de vigilancia.
- Protección contra incendios y sistemas de detección de humo.
- Respaldo de datos y sistemas de recuperación ante desastres.
- Control de acceso a áreas sensibles o críticas mediante la implementación de zonas de seguridad.

Gestión de usuarios

- Autenticación de usuarios mediante contraseñas seguras, autenticación de dos factores o biometría.
- Control de acceso a los sistemas y datos basado en roles y privilegios.
- Implementación de firewalls y sistemas de detección y prevención de intrusiones en los sistemas.

- Monitoreo y registro de eventos de seguridad para detectar y responder a actividades sospechosas.

Controles de red y sistemas

- Firewalls y sistemas de detección y prevención de intrusiones (IDS/IPS).
- Seguridad de la red, como encriptación, segmentación y filtrado de paquetes.
- Actualizaciones y parches regulares del software y sistemas operativos.

Controles de respaldo y recuperación de datos

- Realización de copias de seguridad periódicas de los datos y sistemas críticos.
- Almacenamiento seguro de las copias de seguridad, preferiblemente en ubicaciones externas.
- Pruebas regulares de restauración de datos para garantizar la eficacia de los procedimientos de recuperación.
- Desarrollo de planes de continuidad de negocio y planes de recuperación ante desastres.

Controles organizativos

- Elaboración y aplicación de políticas, procedimientos y normas de seguridad de la información.
- Implementación de procesos de gestión de incidentes y respuesta a emergencias o desastres.
- Evaluaciones regulares de riesgos y revisiones de seguridad.
- Establecimiento de acuerdos de confidencialidad y cláusulas de seguridad en los contratos.

Formación y concienciación

- Programas de concienciación sobre seguridad de la información para todos los empleados.
- Capacitación específica para roles con acceso privilegiado a información con datos personales.

- Pruebas de phishing y simulacros de ataques para mejorar la preparación del personal con acceso a la información.
- Promoción de una cultura de seguridad y fomento de la responsabilidad individual en la protección de la información.

Gestión de incidentes

- Equipo de respuesta a incidentes (ERT) responsable de coordinar y ejecutar las acciones necesarias para responder y recuperarse de los incidentes.
- Plan de respuesta a incidentes donde se describen los pasos y las responsabilidades específicas a seguir durante un incidente.
- Procedimientos de notificación que establecen cómo y a quién se deben informar los incidentes de seguridad.
- Análisis de causa raíz con el objetivo de investigar y determinar las causas fundamentales que llevaron al incidente.

Criptografía

- Uso de algoritmos de cifrado para proteger los datos almacenados y transmitidos por el sistema.
- Gestión de claves criptográficas, incluyendo su generación, almacenamiento y distribución segura.
- Implementación de firmas digitales para verificar la autenticidad e integridad de los datos.
- Protección de datos sensibles mediante técnicas de cifrado adecuadas.

Cumplimiento

- Evaluación y adopción de marcos de cumplimiento, como ISO 27001, D.O.R.A. o PCI DSS.
- Implementación de controles específicos para cumplir con las regulaciones de privacidad de datos, como el RGPD.
- Evaluación y selección de proveedores en base a los criterios de seguridad de la información.

- Establecimiento de acuerdos de seguridad y a nivel de servicio (SLA) con los proveedores.
- Monitorización continua de los proveedores para detectar y mitigar los riesgos de seguridad.

Auditoría

- Implementación de sistemas de registro y monitorización de eventos de seguridad en los sistemas.
- Análisis periódico de los registros de seguridad para identificar posibles brechas o anomalías.
- Realización de auditorías de seguridad internas y externas para evaluar el cumplimiento de los controles de seguridad.
- Realización de auditorías de seguridad a los proveedores para verificar su cumplimiento.
- Implementación de mecanismos de reporte y respuesta a incidentes de seguridad en los sistemas.

4.4 CONSIDERACIONES ÉTICAS Y LEGALES EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL

4.4.1 Reflexión sobre los aspectos éticos relacionados con la seguridad de la información en sistemas de IA

Es de obligada necesidad tener en cuenta que los sistemas de inteligencia artificial están cada vez más presentes en la sociedad y son capaces de tomar decisiones con sus respectivas implicaciones. Por lo tanto, durante todo el ciclo de vida de un sistema de inteligencia artificial se deben definir, al menos, los siguientes principios éticos (tomando como referencia las tres construcciones globales más recientes desarrolladas por la Unesco, la Unión Europea y los principios adoptados por la OCDE - Organización para la Cooperación y el Desarrollo Económicos) [11]:

- **Equidad**: Se debe desarrollar un sistema que favorezca la inclusión y la no discriminación de los usuarios.
- **Transparencia**: Se debe desarrollar un sistema que sea entendible para los usuarios y fomente la confianza en los resultados.
- **Privacidad**: Se debe desarrollar un sistema que fomente el comportamiento responsable de los datos, velando por la protección de los datos personales y la privacidad de los interesados.
- **Responsabilidad**: Se debe desarrollar un sistema que se encuentre supervisado por los humanos con el objetivo de garantizar su correcto funcionamiento, asumiendo la responsabilidad del mismo.
- **Seguridad**: Se debe desarrollar un sistema que se encuentre libre de vulnerabilidades, para esto es necesario conocer las amenazas e implementar las medidas de seguridad oportunas.

En este sentido, y con el foco puesto en la parte ética de los sistemas de inteligencia artificial, existen en el mercado diferentes marcos éticos con el objetivo de poder definir las bases y las líneas de acción a seguir.

A continuación, se describen los marcos más relevantes que existen en la actualidad y son de implicación directa para nuestro país:

Marco Ético para la Inteligencia Artificial, la Robótica y las Tecnologías Conexas de la Unión Europea

Este marco ético establece una serie de principios y pautas para el desarrollo y despliegue ético de la IA en la Unión Europea. Se centra en la transparencia, la responsabilidad, la justicia y la inclusión en el diseño y uso de sistemas de inteligencia artificial [12].

Marco Ético para la Inteligencia Artificial del Gobierno de España

El Gobierno de España ha elaborado un Marco Ético para la inteligencia artificial, que establece los principios éticos y las directrices para el desarrollo y uso responsable de la inteligencia artificial en el país. Este marco busca garantizar la transparencia, la responsabilidad, la justicia y la privacidad en el desarrollo y despliegue de sistemas de inteligencia artificial [13].

4.4.2 Descripción de las regulaciones específicas que pueden afectar a los sistemas de inteligencia artificial.

Ley de Inteligencia Artificial (AI Act)

El AI Act se trata de una propuesta de Reglamento elaborada por la Comisión Europea en abril de 2021**,** como parte de la estrategia digital de la Unión Europea, dónde se quiere regular todo lo relacionado con los sistemas de inteligencia artificial de forma que se asegure el desarrollo y el uso de esta tecnología [14].

En este sentido, y una vez analizada la propuesta elaborada por la Comisión Europea, los objetivos que se esperan con la implantación del AI Act (fecha estimada de implantación para 2025) son los siguientes:

- Abordar los riesgos creados específicamente por las aplicaciones o sistemas de inteligencia artificial.
- Proponer una lista de aplicaciones de alto riesgo.

- Establecer requisitos claros para los sistemas de inteligencia artificial considerados de alto riesgo.
- Definir obligaciones específicas para los usuarios y proveedores de aplicaciones de alto riesgo.
- Proponer una evaluación de la conformidad antes de que el sistema de inteligencia artificial se ponga en servicio o se comercialice.
- Proponer el cumplimiento de la normativa después de que dicho sistema de inteligencia artificial se ponga en servicio o se comercialice.
- Proponer una estructura de gobierno a nivel europeo y nacional para los sistemas de inteligencia artificial.

Reglamento General de Protección de Datos (RGPD) de la Unión Europea

El RGPD establece los requisitos para la protección de datos personales en la Unión Europea. Se aplica a cualquier organización que procese datos personales de ciudadanos de la UE, incluidos los sistemas de inteligencia artificial que utilicen datos personales. El RGPD establece principios de transparencia, consentimiento, minimización de datos y responsabilidad en el tratamiento de datos personales [15].

Directiva de Derechos de Autor de la Unión Europea (Directiva de Copyright)

Esta directiva tiene como objetivo armonizar las leyes de derechos de autor en la UE y abordar los desafíos planteados por la digitalización y la inteligencia artificial. Contiene disposiciones relacionadas con la responsabilidad de los proveedores de servicios en línea en relación con el contenido cargado por los usuarios y la protección de los derechos de autor en el contexto de la inteligencia artificial y el aprendizaje automático.

Propuesta de Ley de Inteligencia Artificial y Derechos Digitales

En mayo de 2021, el Gobierno de España presentó una propuesta de ley específica sobre inteligencia artificial y derechos digitales. Esta ley tiene como objetivo establecer un marco legal para la regulación de la inteligencia artificial, incluyendo aspectos como la transparencia, la responsabilidad, la no discriminación y la protección de los derechos fundamentales en el contexto de la inteligencia artificial.

4.5 REVISIÓN DE SEGURIDAD DE LA INFORMACIÓN EN UN SISTEMA DE INTELIGENCIA ARTIFICIAL BAJO EL ESTÁNDAR ISO/IEC 27001:2022

4.5.1 Metodología de la Revisión

Con el objetivo de poder realizar una evaluación, en materia de seguridad de la información, sobre un sistema de inteligencia artificial se va a emplear como base la metodología propuesta en el estándar internacional ISO/IEC 27001:2022 [16].

La ISO 27001 propone un estándar de seguridad de la información donde se detallan los requisitos y las buenas prácticas a emplear dentro de las organizaciones para poder obtener un adecuado nivel de madurez en sus activos TI (tecnología de la información).

En este sentido, en el anexo A de la norma, se establecen las mejores prácticas para la gestión de seguridad y los controles de seguridad (93 controles), englobados en 4 temas de seguridad, que deben de ser implantados por las organizaciones que busquen la certificación. Los 4 temas de seguridad cubiertos en el estándar son: Organizativos (37 controles), Personas (8 controles), Físicos (14 controles) y Tecnológicos (34 controles) [17].

Teniendo en cuenta estos aspectos, para llevar a cabo la revisión de seguridad de la información bajo el estándar ISO/IEC 27001:2022 se debe seguir la siguiente metodología [18]:

Establecer el alcance

En esta primera etapa se procede a definir el alcance de la revisión, es decir, qué sistemas y procesos de inteligencia artificial serán evaluados. Esto puede incluir tanto el desarrollo de algoritmos de inteligencia artificial como la infraestructura subyacente utilizada.

Revisar la documentación

Se debe obtener acceso a la documentación formal que existe dentro de la organización y que regula el sistema de inteligencia artificial. Esta documentación se compone de las políticas,

procedimientos y normas asociadas a la seguridad de la información, y deben estar en línea con los requisitos de la norma ISO 27001.

Realizar una evaluación de riesgos

Identifica los riesgos asociados con la inteligencia artificial y cómo podrían afectar la seguridad de la información.

Verificar el cumplimiento de los controles

Revisar si se han implementado los controles de seguridad adecuados para mitigar los riesgos identificados y la eficacia operativa de los mismos. Esto incluye los controles establecidos por la ISO 27001 en su Anexo A.

4.5.2 Revisión del Sistema de Inteligencia Artificial

Alcance de la revisión

El alcance de una revisión de inteligencia Artificial se refiere al ámbito y las áreas específicas que serán evaluadas durante el proceso de revisión del sistema. Al definir el alcance de la revisión de inteligencia artificial, se debe establecer de forma clara los límites y las fronteras del sistema a analizar y evaluar. De esta forma, se asegura la eficiencia y el enfoque de la revisión sobre los sistemas de inteligencia artificial.

Dentro de la definición del alcance de la revisión podemos encontrar los siguientes aspectos clave:

- **Objetivo de la revisión**: El objetivo de la revisión es conocer la madurez de aseguramiento de un sistema de inteligencia artificial en base al entandar ISO/IEC 27001:2022.
- **Componentes de la revisión de seguridad**: Los componentes y los procesos implicados serán todos aquellos que queramos asegurar. Este ejercicio se puede realizar desde dos enfoques.
    - Desde **sistemas**: Una vez seleccionados los sistemas de inteligencia artificial sobre los que se quiera hacer la revisión, se deben identificar los procesos de

negocio que hacen uso de los sistemas de inteligencia artificial dentro del alcance.

- Desde **procesos de negocio**: Una vez seleccionados los procesos de negocio sobre los que se quiera hacer la revisión, se deben identificar los sistemas de inteligencia artificial que dan soporte a la ejecución de los procesos de negocio seleccionados.

Por lo tanto, debemos identificar los sistemas y procesos implicados para definir el alcance de la revisión.

- **Fases del ciclo de vida de la Inteligencia Artificial**: Un aspecto clave es el de identificar en qué momento del ciclo de vida se encuentran los sistemas de inteligencia artificial que van a ser revisados, de forma que la evaluación se pueda adaptar a cada situación en función de su estado.
- **Limitaciones y exclusiones:** Establece claramente las limitaciones y exclusiones de la revisión. Esto implica definir qué aspectos específicos que no serán considerados o no estarán dentro del alcance debido a restricciones de tiempo, recursos o cualquier otra consideración relevante.
- **Contexto organizacional**: Ten en cuenta el contexto organizacional en el que operan los sistemas de inteligencia artificial a revisar dentro de la organización objetivo. Esto implica comprender el propósito de los sistemas de inteligencia artificial dentro de la empresa, su relación con los otros sistemas existentes, los impactos potenciales en la organización, etc.

Documentación

En esta fase se deben mantener reuniones con los diferentes actores involucrados en la revisión, de cara identificar la documentación existente en la compañía en materia de seguridad de la información.

La información que se debe recabar son las políticas, procedimientos, normas u otro tipo de documentación en materia de seguridad de la información que haya sido documentada dentro de la organización. En este sentido, la ISO 27001 exige que esta documentación haya sido

aprobada por los responsables en la materia, se haya comunicado a los trabajadores y que se encuentre en un lugar accesible.

De cara a la revisión, esta documentación es el soporte para poder ir revisando si la situación actual de la compañía se encuentra alineada con lo que ha sido descrito en los documentos y poder conocer el grado de cumplimiento interno de la compañía en términos de seguridad de la información.

Riesgos

Para esta etapa se debe tener en cuenta toda la información recogida hasta el momento (contexto de la organización, sistemas implicados, tecnología, ciclo de vida, documentación interna, etc.) para poder aterrizar de una forma concisa los riesgos a identificar en los sistemas de inteligencia artificial.

En el apartado 4.1 de este proyecto ("4.1 DESAFÍOS Y RIESGOS DE LA SEGURIDAD DE LA INFORMACIÓN EN LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL") se ha realizado una identificación a nivel general de los riesgos en materia de seguridad de la información dentro de los sistemas de inteligencia artificial. Por lo tanto, para nuestra revisión se puede partir de ese primer análisis inicial e ir complementándolo con la información adicional y la documentación interna que se ha obtenido de la organización en las etapas anteriores.

Controles

Esta fase en el la parte más relevante de la revisión, donde debemos ir identificado si los controles definidos dentro de la ISO 27001 han sido implementados, en diseño, dentro de la compañía y su eficacia operativa en adecuada para mitigar todos los riesgos en materia de seguridad de la información.

A continuación, vamos a describir los 4 temas que establece la ISO27001, y que contienen los 93 controles de seguridad, de forma que explicaremos de forma detallada en que va a consistir la revisión de cada uno de los temas establecidos y que es lo que exige un marco de control

como la ISO 27001 para garantizar una adecuada madurez en seguridad de la información en los sistemas de inteligencia artificial.

- **CONTROLES ORGANIZATIVOS**
  - EJEMPLOS DE CONTROL
    - Políticas de seguridad de la información
    - Control de acceso
    - Clasificación de la información
  - OBJETIVO
    - El objetivo de los controles organizativos en materia de seguridad de la información es garantizar que esta sea abordada de manera completa y eficaz en todos los niveles de la organización, de forma que se establezca una correcta estructura para promover la protección y el tratamiento adecuado de la información empleada.
  - REVISIÓN

    Durante la auditoría de seguridad de la información de la ISO, y para los controles que se han dado como ejemplo se espera que la organización cumpla lo siguiente.
    - Las políticas de seguridad que ha sido desarrolladas de forma interna se encuentran aprobadas, comunicadas a los empleados y localizadas en un lugar accesible para cualquier persona.
    - Se han definido los procedimientos necesarios para disponer de un control de acceso de los usuarios a los sistemas. En este sentido, se dispone de procedimientos para el alta, modificación y baja de los usuarios, así como la revisión de estos (usuarios genéricos, usuarios inactivos, permisos de los usuarios, etc.)
    - La información que es tratada dentro de la compañía ha sido valorada de forma adecuada, asignándole medidas de seguridad propias para para nivel de información. Además, esta información ha sido etiquetada de forma que es visible para los usuarios con los permisos adecuados.

- **CONTROLES SOBRE LAS PERSONAS**
  - EJEMPLOS DE CONTROL
    - Concienciación, educación y formación en materia de seguridad de la información.
    - Teletrabajo.
    - Reporte de eventos de seguridad de la información
  - OBJETIVO
    - El objetivo de los controles sobre las personas es asegurar que los empleados, proveedores y otros actores involucrados en la organización actúen de manera segura y responsable con los sistemas y los datos almacenados en estos.
  - REVISIÓN

    Durante la auditoría de seguridad de la información de la ISO, y para los controles que se han dado como ejemplo se espera que la organización cumpla lo siguiente.
    - Los empleados y externos reciben formación, a través de cursos o píldoras informativas, en materia de seguridad de la información al inicio de su relación laboral y de manera continua.
    - Se ha definido de forma específica una política de teletrabajo de manera que se establezcan las condiciones y restricciones existentes en la práctica del mismo.
    - Existen canales habilitados para que el personal de la organización pueda reportar en paso de observar o sospechar acerca de un incidente en materia de seguridad de la información. Se debe concienciar a los usuarios de su responsabilidad acerca de reportar cualquier caso o sospecha de incidente a la mayor brevedad posible.

- **CONTROLES FÍSICOS**
  - EJEMPLOS DE CONTROL

- Accesos físicos
- Mesas y pantallas limpias
- Mantenimiento de los equipos

- OBJETIVO
  - El objetivo de los controles físicos en materia de seguridad de la información es garantizar que los activos físicos que almacenan procesan y tratan datos son seguros. Estos controles se encuentran enfocados a asegurar las infraestructuras físicas de la organización (edificios, instalaciones, equipos, soportes de almacenamiento, etc.) de cara a mitigar el riesgo de robo, daño o destrucción de los datos que se encuentran en los mismos.

- REVISIÓN

  Durante la auditoría de seguridad de la información de la ISO, y para los controles que se han dado como ejemplo se espera que la organización cumpla lo siguiente.
  - Se han definido áreas restringidas en los edificios a los que solo puede acceder personal autorizado. Además, estos edificios cuentas con medidas de seguridad físicas, como Circuitos Cerrados de Televisión (CCTV), Guardias de seguridad, etc.
  - Se han definido unas guías de obligatorio cumplimiento y conocimiento del personal donde se definen las diferentes acciones a realizar para proteger la información que pueda estar localizada en las mesas y en los ordenadore, de cara a evitar el acceso no autorizado, la pérdida o el daño de la información en todas sus localizaciones accesibles.
  - Se han definido mantenimientos periódicos sobre los diferentes activos que tratan información dentro de la organización. Estos se encuentran alineados en tiempo y forma con lo recomendados por los fabricantes, son realizados por personal autorizado y supervisados de manera regular.

- **CONTROLES TECNOLÓGICOS**
  - EJEMPLOS DE CONTROL
    - Copias de seguridad (*Backup*)
    - Seguridad en la red
    - Separación de entornos (desarrollo, test y producción)

  - OBJETIVO
    - El objetivo de los controles tecnológicos en materia de seguridad de la información es garantizar los sistemas de información, las redes y los datos de los ciberdelincuentes, de forma que se asegura la integridad, confidencialidad y la disponibilidad de la información.

  - REVISIÓN

    Durante la auditoría de seguridad de la información de la ISO, y para los controles que se han dado como ejemplo se espera que la organización cumpla lo siguiente.
    - La organización dispone de copias de seguridad sobre los datos, el software y los sistemas empleados. Además, se realizan pruebas de simulación para comprobar el correcto funcionamiento de los mismos en caso de necesitar recuperar los datos almacenados.
    - La organización debe asegurarse de implementar los controles apropiados para garantizar la seguridad en la red, de forma que la información se encuentre gestionada y controlada dentro de los sistemas y aplicaciones de la empresa.
    - Se debe comprobar la existencia de entornos separados y seguros. Por lo tanto, deben existir entornos diferentes al de producción para prevenir problemas y poder identificarlos con anterioridad. Además, los datos empleados en el entorno de producción no se deben usar en los otros entornos existentes, evitando así la falta de confidencialidad sobre los datos personales.

4.5.3 Análisis De Los Resultados Obtenidos

Una vez llevada a cabo la revisión de seguridad de la información, en base a la norma ISO/IEC 27001:2022, sobre un sistema de inteligencia artificial, se pueden extraer las conclusiones obtenidas tras el análisis de los resultados obtenidos, identificando los principales puntos del análisis realizado.

En primer lugar, se ha identificado que es importante para las organizaciones implementar controles de seguridad apropiados para abordar las amenazas y vulnerabilidades existentes en los sistemas. En este sentido, la ISO 27001 proporciona un marco sólido para la gestión de la seguridad de la información, que incluye la identificación de riesgos, la implementación de controles de seguridad de la información, la gestión de incidentes y la mejora continua de la seguridad.

Por otro lado, si bien el estándar de la ISO 27001 establece un marco de control que garantiza la seguridad de la información en la mayor parte del sistema, se deberán incluir los aspectos específicos que deben ser analizados por los propios usuarios y entidades certificadoras encargadas.

Además, es necesario incluir controles específicos como por ejemplo para el aseguramiento de la protección de los datos personales. En este caso el estándar internacional ISO 27701 propone controles más enfocados en la protección de los datos personales tratados por los sistemas de información. Sin embargo, el empleo de cualquiera de estos dos estándares ISO (27001 y 27701) no son suficientes para asegurar el entorno de control para garantizar el cumplimiento del RGPD [19].

Adicionalmente al párrafo anterior, se ha identificado que el estándar de la norma de seguridad de la información ISO 27001 no cubre aspectos propios de ciberseguridad (p.e. pruebas de penetración, *pentesting*, etc.) y, por lo tanto, deberían incluirse revisiones de madurez en materia de ciberseguridad de marcos específicos como por ejemplo NIST *Cybersecurity Framework* (NIST CSF).

Por último, se ha identificado que el estándar de la norma de seguridad de la información ISO 27001 deja fuera del alcance de su revisión aspectos relevantes en los sistemas de inteligencia

artificial, tales como: la ética del algoritmo y del sistema, el algoritmo utilizado, los datos de entrenamiento, etc.

5. CONCLUSIONES Y TRABAJO FUTURO

5.1 CONCLUSIONES

5.1.1 Consecución de Objetivos

Una vez finalizado el TFG, se han cumplido los objetivos tanto principales como parciales definidos al principio del proyecto.

El TFG de "SEGURIDAD DE LA INFORMACIÓN EN SISTEMAS DE INTELIGENCIA ARTIFICIAL" ha permitido comprender la importancia de la seguridad en los sistemas de inteligencia artificial mediante la identificación de las amenazas y riesgos específicos a estos sistemas, para posteriormente, adaptar la norma ISO/IEC 27001:2022 a los requisitos de seguridad de los sistemas de inteligencia artificial y llevar a cabo una auditoría de seguridad práctica.

En conclusión, los hallazgos identificados en este proyecto proporcionan una base sólida para abordar los desafíos actuales y futuros en materia de seguridad de la información en los sistemas de inteligencia artificial.

5.1.2 Aspectos Relevantes

La gestión de la seguridad de la información en los sistemas de inteligencia artificial debe de ser un proceso continuo. El catálogo de las amenazas está en constante evolución, y las organizaciones deben estar preparadas para adaptarse y responder a las nuevas amenazas a medida que van apareciendo. Esto implica la disposición de una actitud proactiva por parte de las empresas, mediante la realización de evaluaciones de seguridad regulares y auditorías para identificar y abordar cualquier vulnerabilidad existente en los sistemas.

Por último, y con el objetivo de implantar de forma correcta la triada de seguridad de la información, se requiere un alto compromiso a nivel organizacional. En este sentido, la alta dirección debe comprometerse con la seguridad de la información y debe proporcionar los recursos necesarios para su implementación.

Sin un apoyo adecuado a nivel de dirección, las medidas de seguridad de la información pueden quedar relegadas, lo que puede resultar en vulnerabilidades y aumentar el riesgo de brechas de seguridad.

Desde el punto de vista humano, es crucial fomentar una cultura de seguridad de información dentro de la organización. Esto incluye la formación regular de los empleados en buenas prácticas de seguridad de la información, la promoción de una actitud de responsabilidad con respecto a la protección de los datos y la implementación de políticas de seguridad claras y comprensibles.

Por el lado técnico, se deben implementar y mantener actualizadas las medidas de protección adecuadas, como los sistemas de cifrado y control de acceso, las herramientas de protección antimalware y las soluciones de respaldo y recuperación de datos.

5.1.3 Problemas encontrados

Algunos de los problemas encontrados, a lo largo de la investigación llevada a cabo, para la elaboración del proyecto son los siguientes:

Complejidad del tema

El tema escogido para la elaboración del proyecto "SEGURIDAD DE LA INFORMACIÓN EN SISTEMAS DE INTELIGENCIA ARTIFICIAL" es un campo complejo y que se encuentra en constante actualización.

Si bien se dispone de experiencia previa en el tema (revisiones de seguridad de la información, inteligencia artificial, protección de datos, *frameworks* de seguridad, etc.), ha sido desafiante comprender los conceptos, las técnicas y los retos asociados al proyecto presentado.

Disponibilidad limitada de datos

La seguridad de la información en sistemas de inteligencia artificial es un campo relativamente nuevo, por lo que encontrar información relevante y de calidad para

respaldar la investigación llevada a cabo para el proyecto ha sido una tarea difícil a la par que desafiante.

Desafíos éticos y legales

El campo de la inteligencia artificial plantea cuestiones éticas y legales, que no son de fácil comprensión para un perfil de estudiante más técnico pero que son necesarias para el correcto cumplimiento de la normativa y la adecuación de la tecnología en inteligencia artificial. Además, en el campo legal no han den definido directrices claras hoy en día.

5.2 TRABAJO FUTURO

A continuación, se describen las reflexiones finales y líneas de investigación futura identificadas durante la elaboración del TFG:

Es importante resaltar la importancia de seguir investigando y desarrollando estrategias y marcos de control en materia de seguridad de la información y específicas para los sistemas de inteligencia artificial. Los avances en este campo son rápidos, y es fundamental estar al tanto de las nuevas amenazas y desafíos que van surgiendo.

Asimismo, de cara a proponer posibles líneas de investigación futura en la materia, se identifica la necesidad del desarrollo de estándares de seguridad específicos en inteligencia artificial o la aplicación de técnicas de auditoría automatizada con el uso de la tecnología de inteligencia artificial.

6. BIBLIOGRAFÍA

[1] Seguridad de la Información ¿Qué es la Tríada CID?, https://www.interbel.es/triada-cid/

[2] El ciclo de vida de la inteligencia artificial: alcance, diseño de modelos y despliegue, https://keyrus.com/sp/es/insights/el-ciclo-de-vida-de-la-inteligencia-artificial-alcance-diseno-de-modelos-y#:~:text=Por%20lo%20general%2C%20todo%20ciclo,etapas%20m%C3%A1s%20complejas%20que%20otras.

[3] ¿QUÉ SON 'ENTRENAMIENTO' E 'INFERENCIA' EN INTELIGENCIA ARTIFICIAL, https://neuroons.com/es/que-son-entrenamiento-e-inferencia-en-inteligencia-artificial/#:~:text=%C2%BFQu%C3%A9%20es%20el%20entrenamiento%20de,de%20manera%20veraz%20o%20falsa.

[4] Artificial Intelligence (A Modern Approach), https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf

[5] ISO/IEC 27001, https://www.iso.org/standard/27001

[6] NIST CIBERSECURITY FRAMEWORK, https://www.nist.gov/cyberframework

[7] NIST CSF, https://www.globalsuitesolutions.com/es/que-es-nist-cibersecurity-framework/

[8] Esquema Nacional de Seguridad, https://portal.mineco.gob.es/es-es/ministerio/estrategias/Paginas/Esquema_Nacional_de_Seguridad.aspx

[9] Enfrentando los riesgos de la inteligencia artificial, https://www.mckinsey.com/capabilities/quantumblack/our-insights/confronting-the-risks-of-artificial-intelligence/es-CL

[10] MAGERIT_v3, https://pilar.ccn-cert.cni.es/index.php/docman/documentos/2-magerit-v3-libro-ii-catalogo-de-elementos/file

[11] Cuarenta y dos países adoptan los Principios de la OCDE sobre Inteligencia Artificial, https://www.oecd.org/espanol/noticias/cuarentaydospaisesadoptanlosprincipiosdelaocdesobreinteligenciaartificial.htm

[12] Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un marco de los aspectos éticos de la inteligencia artificial, la robótica y las tecnologías conexas (2020/2012(INL)), https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_ES.html

[13] Estrategia Nacional de IA, https://portal.mineco.gob.es/es-es/ministerio/areas-prioritarias/Paginas/inteligencia-artificial.aspx

[14] Artificial Intelligence Act, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence?at_campaign=20226-Digital&at_medium=Google_Ads&at_platform=Search&at_creation=Sitelink&at_goal=TR_G&at_advertiser=Webcomm&at_audience=ai%20act&at_topic=Artificial_intelligence_Act&gclid=Cj0KCQjwwISlBhD6ARIsAESAmp77c2-m159Dflgat-AvNwEy70gSNFi-nAmCWUH6B3HJdNQ3gn76oOkaAocEEALw_wcB

[15] General Data Protection Regulation, https://gdpr-info.eu/

[16] Norma ISO 27001, https://normaiso27001.es/fase-2-analisis-del-contexto-de-la-organizacion-y-determinacion-del-alcance/

[17] ISO 27001:2022: ¿qué cambios introdujo el nuevo estándar de seguridad?, https://www.aenor.com/conocenos/sala-de-informacion-aenor/notas-de-prensa/espania-sube-en-el-top-ten-mundial-de-las-certificaciones-iso

[18] Ciclo PDCA de gestión de la ISO 27001, https://www.globalsuitesolutions.com/es/ciclo-pdca-iso-27001/

[19] ISO/IEC 27701:2019(en) Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines, https://www.iso.org/obp/ui/#iso:std:iso-iec:27701:ed-1:v1:en

Notas:

Para la obtención de determinados datos y constatar la información se ha hecho uso de:

Servicio web: https://es.wikipedia.org/

Las traducciones se han llevado a cabo con:

Servicio web: https://www.deepl.com/es/translator

7. GLOSARIO DE ACRÓNIMOS Y TÉRMINOS

- **Dato:** representación simbólica de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades.

  Dato. Last access July 2023. https://es.wikipedia.org/wiki/Dato#:~:text=Un%20dato%20es%20una%20representaci%C3%B3n,hechos%20emp%C3%ADricos%2C%20sucesos%20y%20entidades

- **Metadatos: s**on "datos que hablan acerca de los datos", en el sentido de que describen el contenido de los archivos o la información que estos traen en su interior.

  Qué son los metadatos: definición, tipos y ejemplos. Last access July 2023. https://www.docunecta.com/blog/que-son-los-metadatos#:~:text=Definici%C3%B3n%20de%20metadatos&text=La%20palabra%20%E2%80%9Cmetadatos%E2%80%9D%2C%20por,estos%20traen%20en%20su%20interior

- **Regulación:** acción de regular. Medir, ajustar o computar algo por comparación o deducción.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Reglamento:** Colección ordenada de reglas o preceptos, que por la autoridad competente se da para la ejecución de una ley o para el régimen de una corporación, una dependencia o un servicio.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Directriz:** Instrucción o norma que ha de seguirse en la ejecución de algo.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Normativa:** que fija la norma. Conjunto de normas aplicables a una determinada materia o actividad

Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Seguridad de la información:** conjunto de medidas preventivas y reactivas que permiten resguardar y proteger la información. Dicho de otro modo, son todas aquellas políticas de uso y medidas que afectan al tratamiento de los datos que se utilizan en una organización.

  La Seguridad de la Información. Last access July 2023. https://www.tecon.es/la-seguridad-de-la-informacion/

- **Inteligencia artificial (IA):** Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Tecnología de la información:** proceso que utiliza una combinación de medios y métodos de recopilación, procesamiento y transmisión de datos para obtener nueva información de calidad sobre el estado de un objeto, proceso o fenómeno. El propósito de la tecnología de la información es la producción de información para su análisis por las personas y la toma de decisiones sobre la base de la misma para realizar una acción.

  ¿Qué son las tecnologías de la información? Last access July 2023. https://www.ceupe.com/blog/que-son-las-tecnologias-de-la-informacion.html

- **Sistemas de información:** conjunto ordenado de mecanismos que tienen como fin la administración de datos y de información, de manera que puedan ser recuperados y procesados fácil y rápidamente.

  Sistemas de Información. Last access July 2023. https://eade.gt/course/sistemas-de-informacion/#:~:text=Cuando%20se%20habla%20de%20un,y%20procesados%20f%C3%A1cil%20y%20r%C3%A1pidamente.

- **Seguridad informática:** la práctica de proteger equipos, redes, aplicaciones de software, sistemas críticos y datos de posibles amenazas digitales.

¿Qué es la ciberseguridad? Last access July 2023. https://aws.amazon.com/es/what-is/cybersecurity/#:~:text=La%20ciberseguridad%20es%20la%20pr%C3%A1ctica,cliente%20y%20cumplir%20la%20normativa.

- **Riesgo:** Contingencia o proximidad de un daño.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Amenaza:** acción de amenazar. Dar a entender con actos o palabras que se quiere hacer algún mal a alguien.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Vulnerabilidad:** cualidad de vulnerable. Que puede ser herido o recibir lesión, física o moralmente.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Salvaguarda/control:** Defender, amparar, proteger algo o a alguien. Regulación, manual o automática, sobre un sistema.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Framework (marco de control):** Un entorno de trabajo o marco de trabajo es un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar.

  Framework. Last access July 2023. https://es.wikipedia.org/wiki/Framework

- **Ciberataque:** intentos no deseados de robar, exponer, alterar, inhabilitar o destruir información mediante el acceso no autorizado a los sistemas.

  ¿Qué es un ciberataque? Last access July 2023. https://www.ibm.com/es-es/topics/cyber-attack

- **ISO (**Internacional Organization for Standardization) es la Organización Internacional de Normalización, cuya principal actividad es la elaboración de normas técnicas internacionales

  ¿Que es ISO? Last access July 2023. https://www.fundibeq.org/informacion/infoiso/que-es-iso#:~:text=ISO%20(Internacional%20Organization%20for%20Standardization,elaboraci%C3%B3n%20de%20normas%20t%C3%A9cnicas%20internacionales.

- **NIST:** National Institute of Standards and Technology, (NIST) es una institución estadounidense encargada de velar por la innovación y la competitividad industrial**.**

  ¿Qué es el NIST? Last access July 2023. https://veridas.com/que-es-el-nist/#:~:text=El%20National%20Institute%20of%20Standards,innovaci%C3%B3n%20y%20la%20competitividad%20industrial.

- **Algoritmo:** Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema.

  Real academia Española (RAE). Last access July 2023. https://dle.rae.es/algoritmo?m=form

- **Pentesting:** ataque malicioso simulado contra los sistemas informáticos que se usa para encontrar y verificar posibles vulnerabilidades**.**

  ¿Qué es y en qué consiste el pentesting? Last access July 2023. https://www.tokioschool.com/noticias/pentesting/

- **Autenticación:** es la capacidad de demostrar que un usuario o una aplicación es realmente quién dicha persona o aplicación asegura ser.

  Mecanismos de autenticación para verificar la identidad. Last access July 2023. https://redtrust.com/mecanismos-autenticacion/#:~:text=La%20autenticaci%C3%B3n%20consiste%20en%20la,al%20usuario%20con%20dichas%20credenciales.

- **Ética**: Conjunto de normas morales que rigen la conducta de la persona en cualquier ámbito de la vida.

Real academia española (RAE). Last acces July 2023.

https://dle.rae.es/%C3%A9tico

- **Ciberseguridad**: es la práctica de defender, con tecnologías o prácticas ofensivas, las computadoras, los servidores, los dispositivos móviles, los sistemas electrónicos, las redes y los datos de ataques maliciosos llevados a cabo por cibercriminales.

  Definición de Ciberseguridad, Seguridad Informática y Seguridad de la Información. Last Access July 2023.

  https://www.lisainstitute.com/blogs/blog/diferencia-ciberseguridad-seguridad-informatica-seguridad-informacion

- **Privacidad de datos personales:** conjunto de técnicas jurídicas e informáticas encaminadas a garantizar los derechos de las personas sobre el control de su información personal y sobre la confidencialidad, integridad y disponibilidad de esta.

  Preguntas frecuentes en materia de protección de datos personales. Last access July 2023.

  https://www.chj.es/es-es/ciudadano/Atencionalciudadano/Paginas/Preguntasfrecuentesprotecci%C3%B3ndatospersonales.aspx

APÉNDICE I

CATÁLOGO DE AMENAZAS DEFINIDAS EN "MAGERIT – versión 3.0 Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información"

Desastres naturales

Amenazas asociadas a eventos sin la intervención humana

- Fuego: incendios.
- Daños por agua: inundaciones.
- Desastres naturales: tormenta, terremotos, avalancha, tsunami, etc.

De origen industrial

Amenazas asociadas a eventos accidentales o deliberados con la intervención humana

- Fuego
- Daños por agua
- Desastres industriales
- Contaminación mecánica
- Contaminación electromagnética
- Avería de origen físico o lógico
- Corte del suministro eléctrico
- Condiciones inadecuadas de temperatura o humedad
- Fallo de servicios de comunicaciones
- Interrupción de otros servicios y suministros esenciales
- Degradación de los soportes de almacenamiento de la información
- Emanaciones electromagnéticas

Errores y fallos no intencionados

- Errores de los usuarios
- Errores del administrador
- Errores de monitorización (log)
- Errores de configuración
- Deficiencias en la organización
- Difusión de software dañino
- Errores de [re-]encaminamiento
- Errores de secuencia
- Escapes de información
- Alteración accidental de la información
- Destrucción de información
- Fugas de información
- Vulnerabilidades de los programas (software)
- Errores de mantenimiento / actualización de programas (software)
- Errores de mantenimiento / actualización de equipos (hardware)
- Caída del sistema por agotamiento de recursos
- Pérdida de equipos
- Indisponibilidad del personal

Ataques intencionados

- Manipulación de los registros de actividad (log)
- Manipulación de la configuración
- Suplantación de la identidad del usuario
- Abuso de privilegios de acceso
- Uso no previsto
- Difusión de software dañino
- [Re-]encaminamiento de mensajes
- Alteración de secuencia
- Acceso no autorizado

- Análisis de tráfico
- Repudio
- Interceptación de información (escucha)
- Modificación deliberada de la información
- Destrucción de información
- Divulgación de información
- Manipulación de programas
- Manipulación de los equipos
- Denegación de servicio
- Robo
- Ataque destructivo
- Ocupación enemiga
- Indisponibilidad del personal
- Extorsión
- Ingeniería social (picaresca)

Universidad Rey Juan Carlos

APÉNDICE II

PERMISO DE DISTRIBUCIÓN DE RESULTADOS DEL TFG

Datos del proyecto:

Título: SEGURIDAD DE LA INFORMACIÓN EN SISTEMAS DE INTELIGENCIA ARTIFICIAL
Tutor: María De La Paloma Cáceres García De Marina
Cotutor: No aplica
Autor/es: Alejandro García Mayor
Titulación: Grado en Ingeniería Informática
Fecha de defensa: 20/07/2023

Licencia de distribución:

Licencia del software desarrollado como parte del TFG, entregado a través de la aplicación de TFGs (gestion2.urjc.es/tfg). Marque la opción que corresponda:

X Licencia MIT (https://opensource.org/licenses/mit-license.php)
□ Licencia Apache v2 (http://www.apache.org/licenses/LICENSE-2.0)
□ Licencia GPLv3 (https://www.gnu.org/licenses/gpl-3.0.en.html)
□ Otra (Se deberá adjuntar el texto completo de la licencia)
□ No se concede ningún permiso de distribución.

Licencia de la memoria del TFG entregada a través de la aplicación de TFGs (gestion2.urjc.es/tfg). Marque la opción que corresponda:

X Creative Commons Reconocimiento Internacional 4.0 (https://creativecommons.org/licenses/by/4.0/)
□ Creative Commons Reconocimiento-SinObraDerivada 4.0 Internacional (https://creativecommons.org/licenses/by-nd/4.0/)
□ Creative Commons Reconocimiento-CompartirIgual 4.0 Internacional (https://creativecommons.org/licenses/by-sa/4.0/)
□ Otra (Se deberá adjuntar el texto completo de la licencia)
□ No se concede ningún permiso de distribución.

Permiso de distribución:

El Trabajo de Fin de Grado arriba especificado, ha sido defendido y calificado en la Escuela Técnica Superior de Ingeniería Informática de la Universidad Rey Juan Carlos. El tutor (y cotutor si es que existe) del trabajo y su autor (abajo firmantes) expresan su deseo de distribuir los elementos especificados más arriba según las licencias que se mencionan, y en su caso, que se incluyen como anexo.

Lo que ponen en conocimiento de la Universidad.

En Madrid a 18 de julio de 2023

Fdo.: El Tutor	(Fdo: El Cotutor)	Fdo.: Autor/es
Paloma Cáceres García de Marina Firmado digitalmente por Paloma Cáceres García de Marina Fecha: 2023.07.19 11:00:40 +02'00'		Alejandro García Mayor