Universidad Zaragoza
1542

Trabajo Fin de Grado

IMPLEMENTACIÓN DE SISTEMAS DE INTELIGENCIA ARTIFICIAL PARA LA CIBERDEFENSA EN REDES MILITARES

Juan Carlos Morales Rico

Director académico: Dr. D. Gustavo Alares López

Director militar: Tte. D. Carlos Javier Curra Feijóo

Centro Universitario de la Defensa-Academia General Militar

2025

Agradecimientos

Quisiera agradecer en primer lugar a mi familia, pareja y amigos por todo el apoyo recibido durante mi formación.

Del mismo modo, agradecer al personal de la Compañía de Ciberdefensa del Batallón de Guerra Electrónica I/31, por su apoyo y atención en el periodo de prácticas, en especial al teniente D. Carlos Javier Curra Feijóo, Tutor y Director Militar de este TFG, por su dedicación y orientación.

Por último, al profesor D. Gustavo Alares López, por la atención prestada para la realización de este trabajo.

RESUMEN

En las últimas décadas, el papel de la ciberdefensa está adquiriendo una relevancia cada vez mayor en el ámbito militar, consolidándose como un nuevo dominio. Este Trabajo de Fin de Grado tiene como objetivo implementar y evaluar la integración de modelos de Inteligencia Artificial (IA) en los Centros de Operaciones de Seguridad (SOC) que emplean plataformas de Gestión de Información y Eventos de Seguridad (SIEM), con la finalidad de optimizar la clasificación y priorización de alertas, para así fortalecer la capacidad de respuesta ante ciberamenazas en redes militares.

El trabajo se ha desarrollado usando una metodología de investigación aplicada, combinando una revisión bibliográfica con un diseño e implementación de distintos modelos de integración.

El análisis documental muestra la creciente necesidad de automatizar la gestión de eventos en el ciberespacio, especialmente en defensa, donde la velocidad de reacción puede resultar determinante. Se constata que la aplicación de la IA en la ciberdefensa es un campo emergente con un gran potencial para complementar las capacidades humanas y, al mismo tiempo, una necesidad estratégica para hacer frente a las amenazas avanzadas, incluidas aquellas potenciadas por los propios modelos de lenguaje.

Se ha realizado un análisis de necesidades mediante una matriz DAFO, con el objetivo de identificar los factores internos y externos que influyen en la integración. Para llevar a cabo el despliegue experimental se ha utilizado el entorno SIEM de Wazuh sobre sistemas Ubuntu y Docker Desktop, integrado tanto en modelos locales de IA (Mistral de Ollama) a través de scripts de Python, como en modelos en la nube (Claude) a través de servidores intermedios basados en MCP (Model Context Protocol). Para la selección de las herramientas usadas se ha empleado el método AHP (Analytic Hierarchy Process) y una comparación de prestaciones, mientras que para la evaluación de la viabilidad se ha usado el análisis TELOS.

Finalmente, los resultados muestran que esta integración es técnicamente viable y funcional, evidenciando que la IA puede mejorar la eficiencia operativa de los SOC, complementando la labor de los analistas. Asimismo, se pone de manifiesto la importancia de garantizar la soberanía de los datos y de asegurar la conformidad con los marcos normativos y las guías técnicas aplicables.

PALABRAS CLAVE

Inteligencia Artificial (IA), Ciberdefensa, Centros de Operaciones de Seguridad (COS), Ciberataque, Aprendizaje automático

ABSTRACT

In recent decades, the role of cyber defense has been gaining increasing relevance in the military domain, consolidating itself as a new operational sphere. This Final Degree Project aims to implement and evaluate the integration of Artificial Intelligence (AI) models into Security Operations Centers (SOC) that employ Security Information and Event Management (SIEM) platforms, with the purpose of optimizing alert classification and prioritizacion, thereby strengthening the response capability against cyber threats in military networks.

The Project has been developed using an applied research methodology, combining a bibliographic review with the design and implementation of different integration models.

The documentary análisis highlights the growing need to automate event management in cyberspace, especially within defense environments, where reaction speed can be decisive. It is found that the application of AI in cyber defense is an emerging field with great potential to complement human capabilities and, at the same time, a strategic necessity to counter advanced threats, including those enhanced by language models themselves.

A need analysis was carried out using a SWOT matrix, with the aim of identifying the internal and external factors that influence the integration. For the experimental deployment, the Wazuh SIEM environment was used on Ubuntu and Docker Desktop systems, integrated both with local AI models (Ollama's Mistral model) through Python scripts, and a cloud-based model (Claude) through intermediary servers based on the Model Context Protocol (MCP). The selection of the tools employed was carried out using the Analytic Hierarchy Process (AHP) and a performance comparison, while the TELOS analysis was applied for viability assessment.

Finally, the results show that this integration is technically feasible and functional, demonstrating that AI can enhance the operational efficiency of SOCs while complementing the work of analysts. Likewise, the importance of ensuring data sovereignty and compliance with applicable regulatory frameworks and technical guidelines is underscored.

KEYWORDS

Artificial Intelligence (AI), Cyber Defense, Security Operations Centers (SOC), Cyberattack, Machine Learning

INDICE DE CONTENIDO

INDICE DE FIGURAS

INDICE DE TABLAS

ABREVIATURAS, SIGLAS Y ACRÓNIMOS

AHP: Analytic Hierarchy Process

API: Interfaz de Programación de Aplicaciones

CCN: Centro Criptológico Nacional

CERT: Equipo de Respuestas a Emergencias Informáticas

ENS: Esquema Nacional de Seguridad

IA: Inteligencia Artificial

IDS: Sistemas de Detección de Intrusiones

INCIBE: Instituto Nacional de Ciberseguridad

IPS: Sistemas de Prevención de Intrusiones

MCP: Model Context Protocol

NSM: Monitorización de la Seguridad de la Red

SIEM: Gestor de Información y Eventos de Seguridad

SOC: Centro de Operaciones de Seguridad

SOC-D: Centro de Operaciones de Seguridad Desplegable

UEBA: Análisis de Comportamientos del Usuario y la Entidad

DECLARACIÓN DE USO DE INTELIGENCIA ARTIFICIAL

En el presente Trabajo de Fin de Grado se ha hecho uso de diversas herramientas de IA de forma controlada, responsable y complementaria, sin que en ningún momento hayan sustituido el trabajo intelectual, analítico o experimental del autor. Las herramientas empleadas, su finalidad y las medidas de protección adoptadas se detallan a continuación:

- ChatGPT (modelo GPT-5, desarrollado por OpenAI):
    - Para la revisión de la coherencia argumental, gramatical y estilo académico del documento, mejorando la claridad de determinados párrafos.
    - Para la generación de figuras explicativas de arquitectura (Figura 3, Figura 13, Figura 21).
    - Como apoyo en la compresión de documentos bibliográficos ya recopilados por el autor, permitiendo identificar las secciones más relevantes dentro de los textos consultados, sin sustituir la lectura ni el análisis personal de las fuentes originales.
- Mistral (Ejecutado localmente mediante Ollama):
    - Para el desarrollo experimental del trabajo en un entorno local, procesando alertas simuladas generadas en un entorno de pruebas, garantizando la soberanía de los datos al ejecutarse sin conexión a la nube.
- Claude (Anthropic):
    - Para el desarrollo experimental del trabajo en la nube, procesando alertas simuladas generadas en un entorno de pruebas. A diferencia del modelo anterior, al tratarse de un sistema alojado en servidores externos, las pruebas se realizaron con datos ficticios.

En ningún caso se introdujeron datos sensibles. Las consultas se limitaron a cuestiones generales, técnicas o de redacción, sin transmisión de material clasificado, asegurando la seguridad de la información.

1. INTRODUCCIÓN

El ciberespacio en las últimas décadas está pasando a ser considerado uno de los dominios fundamentales para la estrategia en las políticas de defensa y seguridad de los Estados. De hecho, en el contexto de las operaciones militares, ha pasado a considerarse como el quinto dominio, afectando transversalmente a los otros cuatro: tierra, mar, aire y espacio, ya que cualquier acción en este, condiciona e influye al resto de dominios tradicionales. Este nuevo ámbito es definido como un "entorno dinámico en el que confluyen redes de comunicación y sistemas interconectados, con vulnerabilidades y amenazas que afectan directamente a la soberanía y la seguridad nacional". [1] Y en este marco surge la ciberdefensa, siendo el conjunto de medidas para combatir ataques u operaciones cibernéticos que pongan en peligro la soberanía nacional y la seguridad del Estado. [1]

Dentro del ámbito nacional, la Estrategia de Seguridad Nacional reconoce la ciberseguridad como un ámbito importante y clave para la prevención de amenazas y la defensa de los intereses del Estado. También incide en la importancia de la cooperación público-privada para hacer frente a los incidentes que puedan producirse en las infraestructuras críticas del país. [2] Por otra parte, organismos como el Centro Criptológico Nacional (CCN) y el Instituto Nacional de Ciberseguridad (INCIBE) se posicionan como actores fundamentales dentro de las respuestas que puede proporcionar el Estado. El CCN proporciona su acción en el área de la Administración Pública y la gestión de incidentes a través de su Equipo de Respuestas a Emergencias Informáticas (CERT), mientras que el INCIBE orienta su papel a la prevención, formación y concienciación. Estas instituciones reflejan la creciente prioridad que el ciberespacio ocupa en las políticas de seguridad y defensa en la actualidad.

Paralelamente, la Inteligencia Artificial (IA) se define como la habilidad de las máquinas para aprender y desarrollar algoritmos y datos de forma masiva y tomar decisiones de forma autónoma como lo haría un humano. [3] Es uno de los avances más importantes, ya que representa, según la Estrategia de Inteligencia Artificial 2024, "una de las revoluciones más trascendentales de los últimos tiempos". Y dentro del marco estratégico, tiene una gran capacidad para transformar sectores clave, incluidos los de seguridad y defensa. [4]

Cuando esta se aplica al ciberespacio, permite multiplicar las fuerzas en el entorno militar, ya que dota a los sistemas de seguridad de habilidades predictivas y automatizadas. No obstante, este proceso de transformación en el ámbito de la ciberseguridad tiene importantes retos, ya que la IA tiene una doble vertiente, debido a que potencia la detección de amenazas, pero también ayuda a los ciberdelincuentes a realizar prácticas más sofisticadas y peligrosas. Como señala Zambrano (2024), la IA se ha convertido en "una espada de doble filo en el ámbito de la ciberseguridad" [5]. Esta dualidad supone un reto fundamental: aprovechar las ventajas de los modelos predictivos sin exponerse a sus riesgos.

En este escenario es importante aludir al marco normativo y ético que regula la aplicación de IA al ámbito ciberespacial. Su rápida expansión supera la capacidad de los Estados y organismos para regularla, lo que genera vacíos legales y zonas grises. En consecuencia, la evolución tecnológica e integración de IA en distintas herramientas de ciberseguridad, debe ir unida a la regulación y al uso seguro y legítimo de entornos críticos, haciendo especial hincapié en la protección de datos para entrenar algoritmos.

Del mismo modo, estos datos son un elemento central para la aplicación de algoritmos de aprendizaje automático. La eficacia de un sistema depende en gran medida de la calidad y diversidad de la información que se le proporciona. En el caso del ámbito militar, este aspecto tiene un especial énfasis, debido a que los datos manejados son críticos y sensibles para la seguridad nacional, lo que exige no solo un almacenamiento seguro, sino también un proceso de gestión y filtrado de la información permitiendo una toma de decisión efectiva.

Resulta evidente que la ciberseguridad militar debe evolucionar hacia modelos que integren nuevas tecnologías surgidas de la IA, para poder combatir mejor la evolución de las ciberamenazas. Estas soluciones tienen que ser predictivas y automatizadas, tratando de igualar así la velocidad en la que se producen los ciberataques. La idea fuerza no es reemplazar la ciberdefensa tradicional, sino reforzarla. Los analistas humanos aportan criterio estratégico y experiencia en situaciones complejas, pero los grandes volúmenes de datos que se manejan hacen inviable que el humano no trabaje en combinación con herramientas que le faciliten anticipación y agilidad en escenarios de riesgo.

Dentro de este contexto encontramos este trabajo, cuya motivación surge de la problemática real en los Centros de Operaciones de Seguridad (SOC). Estos centros son los encargados de procesar las alertas de ciberamenazas a diario con la ayuda de Sistemas de Gestión de Información y Eventos de Seguridad[1] (SIEM), muchas de estas amenazas suelen ser falsos positivos, lo que genera, en el caso de que haya miles, una sobrecarga y un retraso en la identificación de las alertas que sí son críticas. En el caso de estudio de la empresa Emergía, se documenta que un SOC promedio procesa unas 11.000 alertas diarias, de las que solo un 17% están automatizadas, obligando a los equipos a ignorar hasta un 28% de ellas. [6]

Este Trabajo de Fin de Grado plantea el estudio de la integración de la IA en el ámbito de la ciberdefensa militar, especialmente en la gestión de incidentes, y reforzar la seguridad de las redes militares. Sentando las bases para consolidar la IA como un aliado clave en la defensa estratégica del ciberespacio.

2. OBJETIVOS Y METODOLOGÍA

2.1. Objetivos y Alcance

El objetivo general de este trabajo es analizar la integración de la IA en el ámbito de la ciberdefensa, centrándose en la gestión de un SOC, con el fin de diseñar y evaluar un modelo de aplicación en sistemas SIEM. Como objetivos específicos que ayudan a desarrollar el objetivo general se establecen:

- Realizar una revisión del estado de la cuestión de la aplicación de la IA en ciberdefensa, desarrollando e identificando los retos más importantes.
- Seleccionar y evaluar las herramientas aplicables a un entorno para la gestión de alertas de un SOC.
- Diseñar una arquitectura y posteriormente un prototipo para la integración de IA en un SIEM, analizando su viabilidad, eficacia y rendimiento.

Con respecto al alcance, este trabajo se limitará a un análisis conceptual y experimental en un entorno de pruebas controlado, con el desarrollo de un prototipo de integración de IA en un SIEM, no pretendiendo llevar a cabo un sistema completo listo para su uso en entornos operativos militares, sino sentar las bases de un modelo que sirva de guía para nuevos proyectos más amplios aplicados a escenarios reales.

[1] Un SIEM es una herramienta utilizada en ciberseguridad que recopila y analiza los eventos de una red integrando funciones de monitorización, correlación de eventos y generación de alertas con el objetivo de detectar amenazas, cumplir con normativas de seguridad y facilitar la gestión de incidentes en los SOC.

2.2. Metodología

Este trabajo fundamenta su metodología en la investigación aplicada de carácter experimental, para la aplicación práctica de soluciones en la integración de IA en entornos de gestión de alertas en el ámbito de la ciberdefensa. El proceso seguido en la metodología asegura una estructura de fases que se retroalimentan entre sí, de manera sistemática, verificable y reproducible.

- Revisión bibliográfica y asesoramiento experto

Esta primera fase consiste en una revisión de artículos científicos y académicos de carácter tecnológico, relacionados principalmente con sistemas SIEM, IA, ciberdefensa y ciberamenazas, así como de informes de organismos especializados (p.ej., publicaciones de defensa, estrategias nacionales de ciberseguridad e IA).

Por otro lado, este trabajo se complementa con asesoramiento de expertos en ciberdefensa, con el fin de contrastar la validez de la información seleccionada y comprender mejor el funcionamiento y enfoque de las tecnologías actuales dentro de este campo.

El objetivo de este apartado es establecer un marco teórico sólido, que permita en la siguiente fase identificar las principales necesidades, limitaciones y retos asociados al campo de estudio.

- Análisis de necesidades y recogida de información

El análisis se centrará en identificar los problemas que enfrentan los SOC en la gestión de alertas y constituyen un obstáculo para la eficacia de los sistemas actuales de monitorización y respuesta.

Con el objetivo de complementar este diagnóstico, se recogerá información mediante encuestas a expertos en ciberdefensa y uso de sistemas SIEM. En estas encuestas los participantes deberán valorar una serie de afirmaciones relacionadas con el uso de sistemas SIEM y la IA, utilizando una escala del 1 al 5, donde 1 representa el menor grado de acuerdo y 5 el mayor, con el objetivo de recoger la percepción de los expertos en las problemáticas de gestión de alertas y sus expectativas respecto al uso de la IA como herramienta de apoyo.

- Justificación técnica y diseño

En esta fase se realizará la selección y justificación técnica de los modelos de IA y plataformas SIEM usados en el trabajo, considerando aspectos técnicos y económicos y argumentando la elección final. Esto se llevará acabo realizando una matriz de prestaciones, para comparar los factores críticos de elección, y el método Analytic Hierarchy Process (AHP) como herramienta de apoyo a la toma de decisiones para comparar de forma objetiva las diferentes alternativas. Este método ha sido implementado con la ayuda de un software específico para ello.

Asimismo, se definirá la arquitectura técnica del prototipo y se explicará el desarrollo y el proceso seguido.

- Implementación experimental en entorno de pruebas

En esta fase se llevará a cabo la implementación del prototipo en un entorno de máquina virtual, evitando riesgos en infraestructuras reales. Con la intención de simular un SOC, se instalarán y configurarán los componentes necesarios sobre los cuales se instalará un SIEM que permitirá la generación y gestión de incidencias en forma de alertas de seguridad.

Sobre este entorno se instalará una plataforma de modelos de lenguaje ya entrenados, utilizando uno de ellos para procesar la información recibida del SIEM y generar respuestas útiles para los analistas. Se instalará un modelo preentrenado, evitando el proceso de entrenamiento desde cero y aprovechando las capacidades de un modelo optimizado y listo para funcionar.

- Evaluación de resultados y análisis de viabilidad

La fase de evaluación comprobará si el prototipo cumple los objetivos planteados y si aporta mejoras en eficiencia y priorización de análisis de eventos de seguridad en los SOC, realizando pruebas básicas de funcionamiento.

Finalmente se hará un análisis de viabilidad, aplicando el método TELOS, evaluando como sus siglas indican, los aspectos técnicos, económicos, legales, operativos y de seguridad de los distintos modelos llevados a cabo.

También se planteará una comparativa de soluciones entre entornos en la nube y locales en materia de gestión de seguridad, con el fin de valorar en que escenarios tendría más sentido esta integración.

- Planificación

Por último, la planificación completa de las fases descritas se ha representado mediante un Diagrama de Gantt, que detalla la secuencia temporal del proyecto. Mediante este cronograma se pudieron controlar los tiempos de ejecución y organizar de manera coherente el trabajo, garantizando la realización que las fases de análisis, desarrollo e implementación se llevaban a cabo de manera correcta en tiempo y forma. El diagrama se incluye en el Anexo I.

3. ANTECEDENTES Y MARCO TEÓRICO

3.1. Ciberdefensa

3.1.1. El ciberespacio como nuevo dominio estratégico

Como se señaló anteriormente, el ciberespacio se ha consolidado como el quinto dominio estratégico, y es, por tanto, un nuevo espacio crítico para el desarrollo de las operaciones militares modernas. Este, presenta características muy particulares debido a su dinamismo, al bajo coste de acceso y al anonimato, lo que lo convierten en un escenario muy asimétrico, donde cualquier actor puede comprometer la seguridad de un Estado e incidir en la conflictividad estratégica. [1] El dominio cibernético se encuentra en permanente disputa, los que significa que para las Fuerzas Armadas es crucial la obtención de capacidades específicas de defensa y respuesta.

- Ciberseguridad vs Ciberdefensa

Aunque a menudo estos dos términos se usen como sinónimos y ambos conceptos tengan la misma raíz, son dos términos que presentan diferencias. La ciberseguridad se aplica principalmente al ámbito civil, empresarial o institucional, proporcionando la protección a sistemas de información. Esta garantiza la confidencialidad, integridad y disponibilidad de los datos frente a ciberataques. Por otro lado, la ciberdefensa se aplica al ámbito de las Fuerzas Armadas y, como ellas, se orienta a la protección de los intereses y la soberanía nacional, pudiendo llevar a cabo tanto medidas defensivas como medidas ofensivas. En este escenario queda expuesto que la ciberdefensa se proyecta en los niveles estratégico, operacional y táctico, con un alcance mucho más amplio que la ciberseguridad. [7]

3.1.2. Amenazas en el ciberespacio: actores, técnicas y casos relevantes

Con el objetivo de aportar una visión más clara de las amenazas que nos podemos encontrar en el ámbito del ciberespacio, se mostrarán los diferentes actores que influyen en la ciberdelincuencia, las técnicas que usan y casos de interés.

3.1.2.1. Actores que influyen en la ciberdelincuencia

Tabla 1 Principales tipos de hackers en el ciberespacio. Fuente: Elaboración propia

Tipo de hackers	Descripción	Motivación	Fuente
Sombrero blanco	Especialistas en seguridad que realizan pruebas de intrusión autorizadas y auditoría para mejorar defensa	Protección, prevención, seguridad ética	[8]
Sombrero negro	Actores maliciosos que vulneran sistemas ilegalmente para obtener beneficio o causar daño	Intereses económicos, sabotaje, espionaje	[8]
Sombrero gris	Identifican vulnerabilidades sin permiso para vender la información o explotarlas antes de informar	Reconocimiento, prestigio, beneficio, causas sociales	[8]
Sombrero azul	Individuos con afán de represalias	Venganza	[9]
Phreakers	Precursores del hacking moderno, especializados en manipular sistemas telefónicos	Curiosidad, fraude, telecomunicaciones	[9]
Script-kiddies / Lamers	Usuarios que tienes escaso conocimiento y utilizan herramientas creadas por otros	Diversión, notoriedad, vandalismo digital	[8]
Novatos	Principiantes en proceso de aprendizaje	Formación	[9]

Estos hackers se reúnen en diferentes grupos, que son los que dan lugar a los actores que influyen en la ciberdelincuencia, como pueden ser:

- Grupos APT (Advanced Persistent Threats): Atacantes altamente organizados con un objetivo estratégico claro y prolongado en el tiempo. Relacionado con frecuencia a los hackers sombrero negro.
- Estados: desarrollan capacidades ofensivas y defensivas en el ciberespacio como parte de su estrategia de seguridad nacional. Relacionado en el caso de la ciberdefensa con hackers sombrero blanco, pero en el caso del área ciberofensiva con hackers de sombrero negro
- Hacktivistas: individuos que utilizan el hacking con motivación ideológica, política o social. Relacionados con frecuencia con hackers sombrero gris o negro.
- Cibercrimen organizado: relacionado con frecuencia con hacker sombrero negro, aunque también puede utilizar script kiddies como fuerza auxiliar (ejecutores de ataques masivos con herramientas preconfiguradas).
- Insiders: personas de una organización que sabotean o filtran información crítica. Relacionado con frecuencia con sombreros negros o con sombreros azules dependiendo de la motivación.
- Empresas privadas de ciberseguridad: relacionadas con sombreros blancos.

3.1.2.2. Principales tipos de acciones maliciosas

- Ingeniería social: [10]

Son el conjunto de técnicas de ciberdelincuencia basadas en el engaño y la manipulación.

  - Phising / Vishing / Smishing: correos, llamadas o mensajes SMS que suplantan la identidad de personas de confianza con la intención de robar información, credenciales o propagar un archivo malicioso.

Figura 1 Ejemplo de phishing [11]

  - Baiting (cebo): el atacante deja un cebo físico o digital para que la víctima interactúe con él y se infecte.

- Ataques a las conexiones: [12]

Conjunto de técnicas basadas en intercambio de información por conexiones inalámbricas entre los usuarios y la infraestructura de red, para monitorizar y robar datos.

  - Redes trampa (puntos Wi-Fi falsos): creación de un punto de acceso que obliga a los usuarios a conectarse, para interceptar tráfico o inyectarlo.
  - Spoofing: suplantación de identidad a nivel de red o servicios, para redirigir usuarios a sitios de phishing o interceptar comunicaciones.
  - Ataque a cookies: mediante técnicas como XSS[2] o Session fixation[3], el atacante obtiene un "token"[4] de sesión que aceptará el servidor como si fuese la propia víctima, es decir, el atacante "secuestra" la sesión sin conocer la contraseña.
  - Denegación de servicio distribuida (DDoS): Saturación de recursos mediante redes de bots o tráfico ilegítimo para deshabilitar los servicios en línea.
  - Inyección SQL: Alteración en la programación de una base de datos (p.e. una base de datos de credenciales) infiltrando código.

[2] XSS o Cross-Site Scripting es "un tipo de ataque en el cual actores maliciosos logran inyectar un script malicioso en un sitio web". Fuente: https://itequia.com/es/jwt-o-cookies-de-sesion-autenticacion-entornos-web/

[3] Tipo de ataque en el que el atacante obtiene un "token" de sesión válido y este induce a que la víctima lo use. Fuente: https://owasp.org/www-community/attacks/Session_fixation

[4] Un "token" es un identificador digital único generado por el servidor que autentica a un usuario en un sistema web y permite mantener la sesión activa sin enviar repetidamente las credenciales. Fuente: https://itequia.com/es/jwt-o-cookies-de-sesion-autenticacion-entornos-web/

- Escaneo de puertos: técnica para descubrir puertas abiertas o puntos débiles en una red.
- Man-in-the-Middle y sniffing: interposición en una comunicación para capturar o modificar servicios.

- Ataques por malware: [10][12]

Programas maliciosos con el objetivo de producir efectos dañinos en la víctima, como robo, extorsión y destrucción.

- Virus: código que se inserta en ficheros y se reproduce, alterando el comportamiento esperado de sistemas y aplicaciones.
- Adware / Spyware: software que tiene como objetivo la inserción de publicidad intrusiva.
- Troyanos: programas camuflados que abren puertas traseras o roban credenciales.
- Rootkits: técnicas que permiten a los atacantes tomar el control de las máquinas y robar datos.
- Ransomware: malware que cifra y secuestra datos con la intención de extorsionar.

Figura 2 Ejemplo de ransomware [13]

Diferentes actores y técnicas de ciberdelincuencia han tenido especial repercusión en casos relevantes que se van a enumerar a continuación:

Tabla 2 Casos relevantes de ciberataques. Fuente: Elaboración propia

Caso	*Actores*	*Técnica*	*Descripción*	*Respuesta*	*Fuente*
Ciberataques a Estonia (2007)	*Hacktivistas prorrusos*	*DDoS masivos*	*Durante 22 días se lanzaron ataques contra bancos, medios y organizaciones, paralizando la infraestructura digital*	*El gobierno de Estonia reforzó sus defensas nacionales y solicitó apoyo a la OTAN y a la UE*	*[14]*

Ataque a la red satelital KA-SAT (Viasat 2022)	*Atacantes vinculados a Estados, en este caso Rusia*	*Malware wiper*[5]	*Se inutilizaron 40.000 módems en Europa, afectando comunicaciones militares ucranianas.*	*Viasat reemplazó miles de módems e implementó protocolos de seguridad reforzados.*	*[15]*
Operaciones del Grupo Lazarus (2009-actualidad)	*Grupo APT norcoreano*	*Phising, ransomware*	*Son los responsables del hackeo a Sony Pictures (2014), robo a instituciones como el Banco Central de Bangladesh (2016) y de la creación del ransomware WannaCry, usando el exploit*[6] *EternalBlue*[7] *para aprovechar una vulnerabilidad de Windows*	*Tras WannaCry, GCHQ británico y Microsoft publicaron parches de emergencia*	*[16]*

3.1.3. La ciberdefensa militar

- La ciberdefensa en el nivel estratégico

La ciberdefensa es un sistema complejo en el que intervienen múltiples actores militares, civiles y privados, que se integra en la planificación nacional y requiere una coordinación internacional. [17] Organizaciones como la OTAN y la Unión Europea ya impulsan este dominio con la creación de la ENISA[8] y la NATO CCDCOE[9] respectivamente. Además, la Alianza Atlántica promueve ejercicios multinacionales como el Locked Shield[10].

El enfoque de defensa en el ciberespacio resulta prioritario en los roles críticos que tienen las rutas y puertos comerciales, responsables de más del 90% del comercio global y objetivo primordial de los ciberdelincuentes. Por tanto, la cooperación entre aliados y socios y las relaciones internacionales en el ámbito estratégico de la ciberdefensa garantizan el fortalecimiento de la ciberdefensa a escala global.

- La ciberdefensa en el nivel táctico y operacional

Las operaciones a menos escala, es decir las aplicadas directamente sobre las unidades y operaciones militares, son la integración de capacidades de protección de redes y guerra electrónica en brigadas y batallones que aseguran la superioridad en el espectro electromagnético. [18]

Dentro de los SOC desplegables (SOC-D) esta aplicación se hace efectiva en el ciberespacio por medio de los Sistemas Integrados de Monitorización Búsqueda y Análisis

[5] Los wipers constituyen una subcategoría de malware destructivo que elimina o destruye el acceso de una organización a archivos y datos. Fuente: https://www.checkpoint.com/es/cyber-hub/threat-prevention/what-is-malware/what-is-wiper-malware/

[6] Un exploit es un programa que aprovecha vulnerabilidades para provocar comportamientos como la toma de control de un sistema. Fuente: https://www.pandasecurity.com/es/security-info/exploit/

[7] EternalBlue fue un exploit creado por la Agencia de Seguridad Nacional de los EE. UU. como parte de un controvertido programa que tenía el objetivo de almacenar y armarse con vulnerabilidades de ciberseguridad. Fuente: https://www.avast.com/es-es/c-eternalblue

[8] La Agencia de la Unión Europea para la Ciberseguridad (ENISA) es el organismo de la UE encargado de apoyar las políticas de ciberseguridad. Fuente: https://www.enisa.europa.eu

[9] El NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) es un centro de excelencia acreditado por la OTAN dedicado a la investigación, formación y ejercicios en ciberdefensa. Fuente: https://ccdcoe.org

[10] Locked Shield es el mayor ejercicio internacional de ciberdefensa del mundo, organizado por el NATO CCDCOE para entrenar y evaluar la capacidad de respuesta frente a ciberataques. Fuente: https://ccdcoe.org/exercises/locked-shields

(SIMBA). Estos son sistemas integrales, modulares, portables y autónomos que permiten implementar las capacidades de ciberdefensa en cualquier escenario de un teatro de operaciones. Se componen de 5 módulos para cubrir las distintas capacidades:

- Módulo de monitorización: Su función es vigilar en tiempo real las redes, sistemas y servicios para detectar actividades anómalas o sospechosas mediante los SIEM. Actúa como primera línea de defensa, ya que detecta y genera alertas en el mismo momento en que se produce un ataque. El medio hardware principal en este módulo son las sondas, servidores donde están instalados los softwares de los SIEM.
- Módulo de inspección y pentesting: Se encarga de revisar y poner a prueba la seguridad de los sistemas y aplicaciones haciendo uso de auditorías de seguridad y pruebas de penetración (pentesting) simulando ataques reales. El objetivo no es comprometer el sistema, sino identificar puntos débiles y proponer medidas. Los medios hardware de este módulo son dispositivos que permiten simular intrusiones, como memorias USB programadas para realizar ataques rápido o dispositivos que emula puntos de acceso WiFi maliciosos.
- Módulo de análisis malware: encargado de estudiar programas utilizados en campañas de ataque. Se realizan técnicas de ingeniería inversa, sandboxing[11] y monitorización de progresos, para obtener información acerca de cómo se propaga y como ataca, así como medir los daños que puede causar. Este apartado contribuye a dos importantes partes, tanto a la defensa operativa como la inteligencia de amenazas.
- Módulo forense: se realizan acciones de investigación para reconstruir los hechos tras un incidente. Se utilizan medios para la recolección y preservación de evidencias digitales, así como medios clonadores de dispositivos, y se realiza un análisis cronológico de eventos.
- Módulo de gestión: En este módulo se manejan plataformas de gestión de incidentes y ticketing[12], herramientas de análisis de riesgos y métricas de seguridad y se realizan informes y gestión documental de los procedimientos e incidentes ocurridos.

3.1.4. IDS, IPS y SIEM: funciones y complementariedad en los SOC

En el ecosistema de la monitorización los sistemas de detección y prevención de intrusos (IDS e IPS) representan la primera línea de defensa frente a ataques, mientras que los SIEM complementan esta arquitectura. La comprensión de sus diferencias y su complementariedad son esenciales para entender los SOC modernos.

Los IDS (Intrusion Detection System) son sistemas pasivos cuya misión es detectar actividades anómalas en redes y equipos, analizando tráfico y registrando los sistemas buscando patrones sospechosos. [19] Estos, al ser pasivos, no bloquean los ataques, sino que se encargan de avisar a los administradores o sistemas superiores.

Los sistemas que si se encargan de intervenir de manera activa bloqueando y mitigando ataques en tiempo real son los IPS (Intrusion Prevention System). [19] De esta forma se impide

[11] El sandboxing es una práctica de ciberseguridad que consiste en ejecutar código malicioso en un entorno controlado y aislado de la red, para observarlo y analizarlo. Fuente: https://www.checkpoint.com/es/cyber-hub/threat-prevention/what-is-sandboxing/

[12] Las herramientas de ticketing son software que permiten gestionar solicitudes, incidencias o consultas de clientes ofreciendo ticket para controlar el seguimiento de estas. Fuente: https://www.ringover.es/blog/ticketing-herramientas

que el ataque llegue a los sistemas críticos, ya que actúa en línea con el tráfico. Por esta razón, los sistemas IPS pueden interrumpir el servicio y cortar el tráfico, algo que puede ser positivo para la infraestructura. Sin embargo, también puede resultar perjudicial si se usa en sistemas donde un bloqueo interrumpe procesos críticos, algo que no ocurre con los IDS, ya que aportan una visibilidad sin riesgos.

Aunque estos dos sistemas son fundamentales para detectar y detener intrusiones, generan alertas y logs en cantidades difíciles de gestionar si se tratan de entornos complejos. Los SIEM integran estas funcionalidades para la correlación y análisis de los datos que provienen de distintas fuentes, incluyendo IDS, alimentando a la plataforma con datos de intrusiones, e IPS, aportando la parte de respuesta automática.

Las plataformas SIEM, constituyen la columna vertebral de la mayoría de SOC. Estos sistemas proporcionan una visión centralizada del estado de seguridad de una organización. [19] En conjunto, el SOC gana capacidad de detección profunda, de prevención inmediata y de análisis estratégico.

El funcionamiento de un SIEM se articula en varias fases encadenadas. Primero se produce la ingesta de datos, en la que se recopilan datos de diferentes fuentes a través de conectores, agentes y protocolos. Estos datos pueden ser logs, eventos y registros de seguridad de fuentes como firewalls, IDS e IPS, endpoints, aplicaciones, etc. Tras esto, los datos en bruto pasan al indexador, aquí se normalizan y clasifican para almacenarse en bases de datos. Después el SIEM inicia la fase de análisis y correlación, aplicando reglas predefinidas a eventos para detectar patrones y comportamientos sospechosos. La supervisión de estos datos da como resultados alertas priorizadas según su criticidad, que se visualizan en paneles o dashboards, donde los analistas monitorizan de manera centralizada las incidencias.

Figura 3 Proceso de un SIEM. Fuente: ChatGPT

3.1.5. Retos y perspectivas de la ciberdefensa militar

El desarrollo de la ciberdefensa en redes militares encara desafíos y retos, que lejos de estar en un futuro lejano, están ya muy próximos. La rápida evolución de las amenazas y de los entornos híbridos, junto con la característica transversal del ciberespacio y la dificultad de atribución de los ciberataques, limitan de manera excepcional la capacidad de respuesta diplomática y militar.

La dependencia tecnológica del mundo actual supone una preocupación que deriva en una necesidad de resiliencia y redundancia en los medios, garantizando la continuidad operativa en caso de ataque.

La integración en la doctrina militar con la transformación digital de las Fuerzas Armadas, como refleja el proyecto de Fuerza 35, supone otro esfuerzo adicional, tratando de contemplar este ámbito como parte de una maniobra conjunta, en lugar de como un ámbito aislado. [18]

Por último, como reto formativo, se debe incluir el adiestramiento especializado que suponen las capacidades de ciberdefensa, con el entrenamiento de las Fuerzas Armadas en escenarios realistas y ejercicios multinacionales, así como la inversión que esto supone.

3.2. Inteligencia Artificial

3.2.1. Definición y fundamentos

La IA es hoy en día unos de los pilares del cambio tecnológico y social contemporáneo. Podemos definir esta como el conjunto de sistemas computacionales que tienen la capacidad de emular las funciones cognitivas humanas, tales como el razonamiento, la toma de decisiones y el aprendizaje. [20] De manera más simplificada Rouhiainen (2018) define la IA como “la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana”. [3]

Los fundamentos de la IA se basan en tres pilares, el primer pilar son los algoritmos que permiten el aprendizaje autónomo, el segundo pilar son los datos masivos, los cuales alimentan a los modelos de IA y, por último, la capacidad computacional.

En un primer momento, los desarrollos de IA se centraron en IA basada en conocimiento, reproduciendo reglas con el fin de reproducir el razonamiento de los humanos en un dominio en concreto. En este caso, el sistema siguiendo una serie de instrucciones explícitas resuelve el problema, por tanto, su adaptación en entornos cambiantes es poco eficaz, provocando un declive de esta forma de IA simbólica.

Desde los años 2000, un nuevo paradigma se está consolidando: la IA basada en datos, basada principalmente en aprendizaje automático en un primer momento y aprendizaje profundo más recientemente, con la premisa de extraer patrones a ingentes cantidades de datos mediante algoritmos estadísticos de aprendizaje. Estos patrones son los que sirven para aportar las soluciones a los problemas planteados. [3]

En la actualidad, los sistemas o modelos que utilizan tanto aprendizaje automático como aprendizaje profundo son referidos como IA.

El aprendizaje automático o machine learning, se lleva utilizando décadas para la personalización de medio sociales como Facebook o motores de búsqueda como Google, y es un aspecto de la informática que explota los algoritmos más allá y hace que los ordenadores tengan la capacidad de aprender sin estar programados para ello.

Tipos de aprendizaje automático: [3]

- Aprendizaje supervisado: el sistema se entrena con datos que han sido etiquetados anteriormente, como por ejemplo con correos clasificados como “spam” o “no spam”. Se utiliza mayoritariamente en aplicaciones comerciales y requiere la intervención humana para poder proporcionar retroalimentación.
- Aprendizaje no supervisado: el sistema etiqueta datos de manera autónoma, sin necesidad de ayuda y agrupándolos él mismo.
- Aprendizaje por refuerzo: los algoritmos aprenden basándose en la experiencia, y cada vez que aciertan hay que darle un refuerzo positivo, una recompensa.

Por otra parte, tenemos el aprendizaje profundo o deep learning, que se trata de un subcampo del aprendizaje automático y hace uso de redes neuronales para resolver problemas muy complejos con grandes cantidades de datos y de potencia de procesamiento. En palabras más sencillas, es una “evolución” del aprendizaje automático, y es la base de los principales modelos de IA.

3.2.2. Riesgos y desafíos

El avance de la IA conlleva amenazas a la evolución técnica, política y social de la humanidad. Por un lado, uno de los riesgos identificados es la dependencia estratégica. Las infraestructuras de supercomputación capaces de manejar los enormes volúmenes de datos que requieren los grandes modelos están concentradas en pocas potencias mundiales y grandes corporaciones tecnológicas. Esto genera una dependencia peligrosa y plantea interrogantes alrededor de la soberanía digital y la autonomía de la tecnología de los Estados. [4]

Por otro lado, la IA, al igual que ayuda a luchar contra la ciberdelincuencia y aporta mucha fuerza en el ámbito de la ciberdefensa como se verá más adelante, también ayuda a los actores maliciosos a perfeccionar sus tácticas. Como ejemplos de cómo se utiliza la IA para realizar ciberataques podemos ver: [5]

- Generación masiva automática de noticias falsas o videos y audios falsos para manipular la opinión pública.
- Robo y reutilización de contenido digital, algoritmos de IA clasifican y ordenan informaciones masivas para realizar fraudes o campañas de extorsión.
- Creación de bots inteligentes para realizar ataques DDoS más avanzados, teniendo la capacidad de modificar patrones de ataque en tiempo real, en lugar de generar tráfico indiscriminadamente. Los bots incluso imitan comportamientos humanos y realizan interacciones falsas dificultando su detección.
- Phising asistido por IA, con visión artificial capaz de crear páginas web falsas indistinguibles, e incluso vishing (phishing por voz) clonando acentos y tonos de personas reales.
- Acceso de script kiddies a capacidades avanzadas, este tipo de hackers con la IA tiene la capacidad de lanzar ataques más complejos, haciéndose ayudar por generadores de malware básico, o instrucciones detalladas para realizar ataques automatizados todo ello basado en IA. Esto significa que personas sin experiencia técnica, pueden desarrollar herramientas maliciosas con un nivel de complejidad elevado, multiplicando el número potencial de atacantes y la peligrosidad del ciberespacio.

3.3. Inteligencia Artificial aplicada a Ciberdefensa

3.3.1. Integración de IA en SIEM

Los SIEM tradicionalmente se han basado en reglas estáticas y correlaciones predefinidas, lo que resulta poco eficaz en escenarios limitados, por tanto, se plantean dos problemas:

- Incapacidad para detectar ataques novedosos de actores sofisticados o con gran disponibilidad de recursos.
- La sobrecarga de alertas y falsos positivos, que saturan a los analistas y dificulta la gestión eficaz de los incidentes.

Aquí es donde la IA amplía las capacidades de los SIEM permitiendo la detección de anomalías basadas en comportamientos, más allá de reglas predefinidas y realizando predicciones de los vectores de ataques futuros a partir de patrones históricos.

Sin embargo, el problema concreto sobre el que este proyecto pretende centrarse es el segundo punto, la gestión de alertas. La carga excesiva de alertas limita la eficacia operativa de los SOC y retrasa la identificación efectiva de incidentes. La incorporación de técnicas de machine learning y deep learning en las plataformas SIEM para solucionar este punto son:

- La capacidad de priorizar de manera inteligente alertas, evaluando la criticidad de cada alerta en función del contexto y el posible impacto que puede tener en activos críticos, permitiendo asi que los analistas centren sus esfuerzos en incidencias importantes.

- Reducción de falsos positivos, debido a que los sistemas con integración IA pueden identificar los eventos que anteriormente han sido catalogados como benignos y ajustar las alertas para no generar incidentes innecesarios.
- Visualización avanzada: con la capacidad de análisis de grandes volúmenes de datos, los modelos de aprendizaje autónomo pueden ofrecer dashboards dinámicos y sistemas de visualización que favorezcan la toma de decisiones veloz en contextos críticos.

3.3.2. Sistemas Actuales de IA integrada en SIEM

El cuadrante de Gartner para SIEM [21] de 2024 confirma la tendencia de uso de módulos de IA para ir más allá de la tradicional correlación de reglas para la generación de alertas. Proveedores como Splunk o Microsoft Sentinel se mantienen líderes en el cuadrante, y otros como Google Security Operations han irrumpido en el panorama como visionarios, gracias en parte a la incorporación de IA generativa (Gemini) aplicada a búsquedas, optimización de reglas y generación de contexto automático para los analistas. Este reconocimiento que sea ha plasmado en Gartner demuestra que la IA se ha convertido en un criterio de diferenciación estratégica en el mercado de SIEM y no solo una característica técnica opcional.

En el panorama actual pueden distinguirse tres enfoques principales. El primero corresponde a los SIEM comerciales con integración nativa de IA, centrándonos en los modelos de pago que van destinados a grandes corporaciones o administraciones públicas. En este ámbito los más representativos son Splunk Enterprise Security e IBM QRadar. Splunk, se ha mantenido líder en el cuadrante de Gartner gracias a la incorporación de módulos UEBA[13] basados en machine learning. Estos módulos priorizan alertas dependiendo del contexto y son capaces de asignar puntaciones de riesgo tanto a usuarios como a equipos y aplicaciones. Con estas puntuaciones los analistas pueden centrarse en las alertas de mayor riesgo y descartar falsos positivos. Por otro lado, IBM QRadar dispone de Watson. Esta IA es capaz de automatizar la gestión de alertas gracias al procesamiento de lenguaje natural y puede filtrar y enriquecer las alertas con información contextual, así como realizar una priorización de alertas inteligente, ayudar en la exportación de resultados o realizar investigaciones cognitivas de incidentes [22].

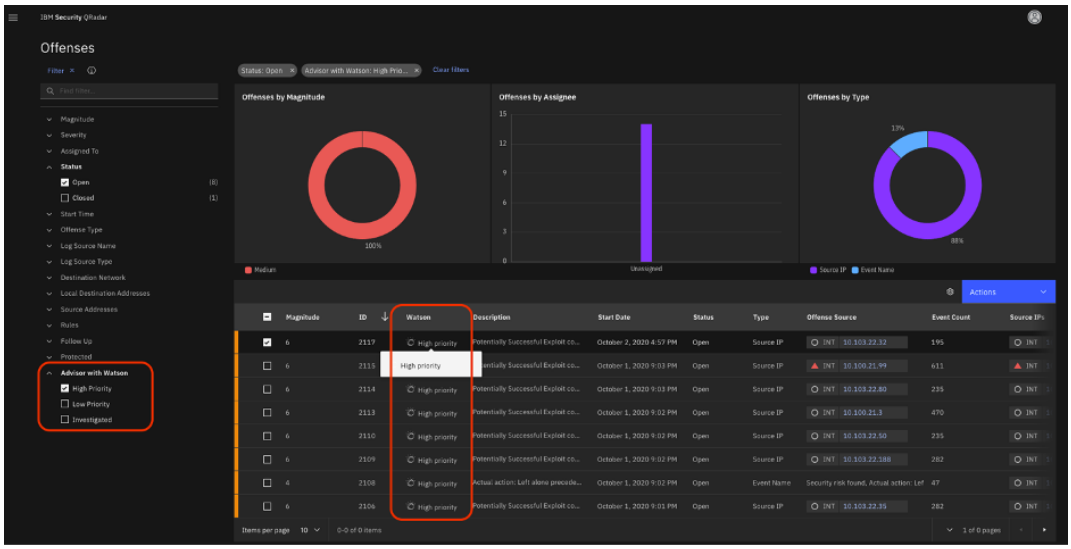


Figura 4 Ejemplo de asignación de prioridades a alertas en IBM QRadar con Watson [23]

[13] El User and Entity Behavior Analytics (UEBA) o Análisis de Comportamientos del Usuario y la Entidad en español, es una solución de ciberseguridad que utiliza algoritmos de aprendizaje automático con los que detectan anomalías en el comportamiento de usuarios y equipos. Fuente: https://www.fortinet.com/lat/resources/cyberglossary/what-is-ueba.

Estos proyectos ilustran que la IA ya no se concibe como un añadido, sino que es un componente nativo, sin embargo, estas soluciones presentan una limitación clara: el elevado coste de licencias y la dependencia de un único proveedor.

El segundo enfoque se encuentra en el ámbito de los SIEM de código abierto o gratuitos con integración de IA, donde, aunque las capacidades no siempre son nativas, permiten incorporar modelos experimentales o basados en la comunidad. Uno de estos sistemas es Security Onion, que combina motores de detección (IDS) como Suricata y Zeek con una función experimental de machine learning llamada AI Summary que reduce la curva de aprendizaje, ya que facilita que los analistas menos experimentados comprendan las reglas avanzadas, acelera la validación de alertas proporcionando contexto sobre el motivo de la incidencia y mejora de esta manera la eficiencia operativa. [24]

Ejemplo de AI Summary, donde convierte una regla de detección en un resumen práctico y explicado en lenguaje natural:

Figura 5 Regla de detección [24]

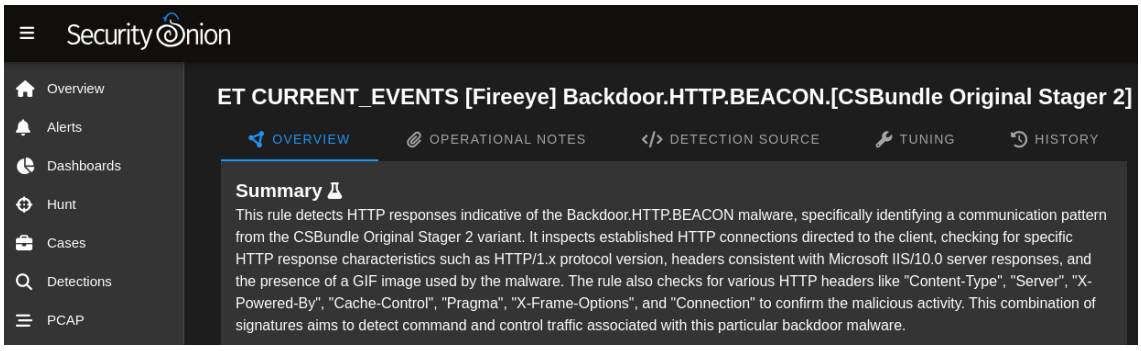


Figura 6 Resumen de AI Summary para la regla de detección [24]

Otro ejemplo de AI Summary, esta vez explicando una regla que detecta la vulnerabilidad CVE-2024-1212 en lenguaje natural:

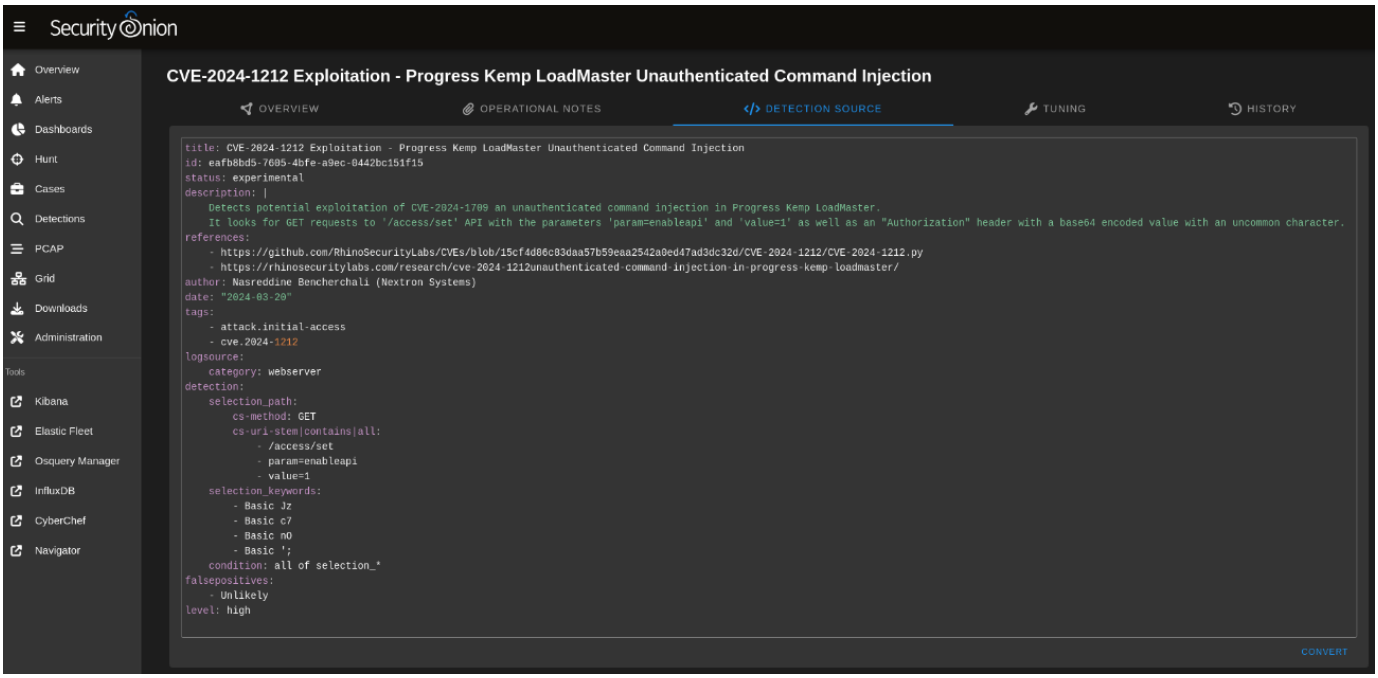


Figura 7 Regla de detección de vulnerabilidad CVE 2024-1212 [24]

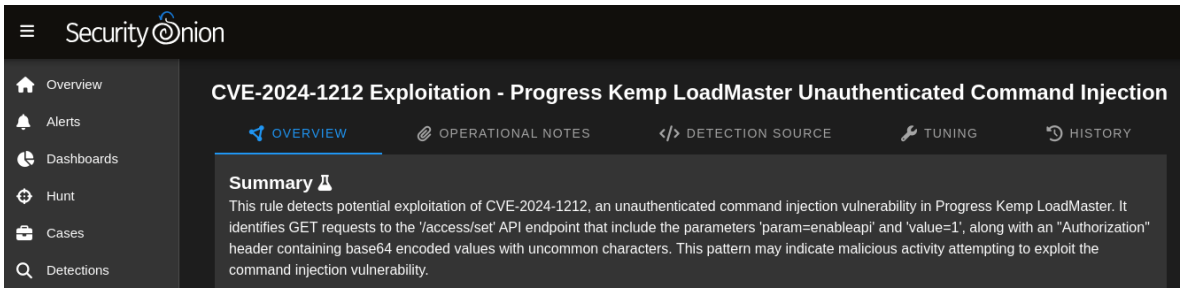


Figura 8 Resumen de AI Summary para la regla de detección de vulnerabilidad CVE 2024-1212 [24]

El tercer enfoque es especialmente relevante para este proyecto, se trata de los SIEM flexibles que integran IA externa mediante Interfaz de Programación de Aplicaciones (API)[14] o protocolos estandarizados. Dentro de este grupo destaca Wazuh, que, aunque no cuenta con capacidades de IA nativas, su arquitectura abierta permite integrar el sistema con modelos de lenguaje como Claude o Mistral. De esta forma, las alertas que genera el SIEM pueden ser enriquecidas con un análisis de contexto y organizadas por prioridad con una clasificación inteligente. El caso de la implementación de Wazuh con un modelo de IA constituye un ejemplo claro de cómo una herramienta gratuita y abierta puede alcanzar capacidades comparables a las soluciones comerciales de alto nivel.

De la misma forma, otras herramientas están siendo desarrolladas en la misma línea que la anterior, los Model Context Protocol (MCP), que establecen estándares de comunicación entre los modelos de IA y otras herramientas externas. Este protocolo funciona como capa intermedia para conectar de manera controlada los modelos de lenguaje con las herramientas y datos del SOC.

[14] Las APIs son mecanismos que permiten a diferentes componentes de software comunicarse entre sí con un conjunto de protocolos y herramientas. Fuente: https://aws.amazon.com/es/what-is/api/#:~:text=en%20su%20tel%C3%A9fono.-,%C2%BFQu%C3%A9%20significa%20API%3F,de%20servicio%20entre%20dos%20aplicaciones.

En esta arquitectura, el SIEM expone una serie de funciones MCP y el modelo de IA en lugar de recibir solo un prompt, invoca estas herramientas de forma estructurada para enriquecer de esta manera su análisis.

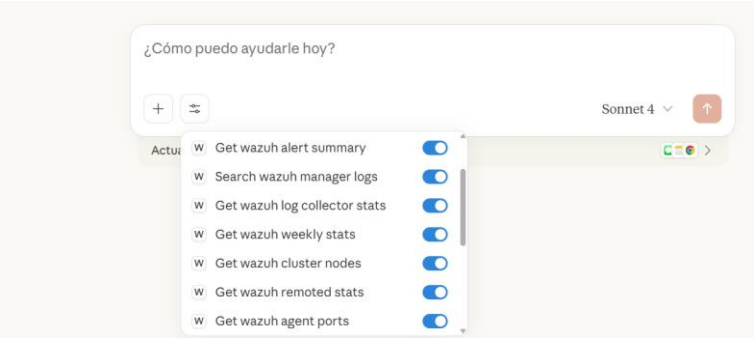


Figura 9 Ejemplo de funciones MCP en el modelo de IA Claude para invocar herramientas de Wazuh. Fuente: Elaboración propia

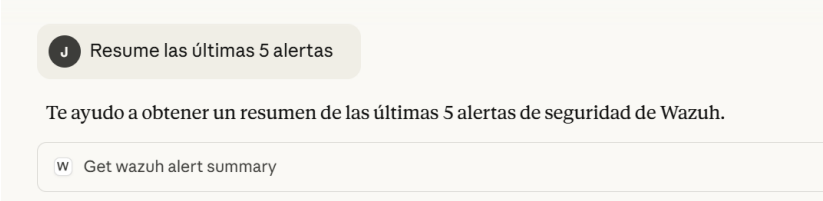


Figura 10 Ejemplo de la invocación de las herramientas de Wazuh en Claude con funciones MCP. Fuente: Elaboración propia

El estado de la cuestión muestra que los diferentes entornos tanto comerciales como de código abierto o entornos híbridos convergen en ofrecer soluciones a las tres funciones críticas de la gestión de alertas: la priorización inteligente, la reducción de falsos positivos y la visualización avanzada. Mientras que la detección se encuentra cada vez más extendida y automatizada, la gestión eficiente de alertas mediante IA tiene un impacto directo en llegar a una ciberdefensa más ágil y precisa para los analistas.

3.3.3. Otros casos de uso de la IA en Ciberdefensa

La incorporación de la IA en la ciberdefensa está transformando la gestión que los Estados y las Fuerzas Armadas llevan a cabo del ciberespacio. La capacidad de predicción de amenazas, de ayuda en la toma de decisiones y de gestión de datos, hace imprescindible la implementación de esta para combatir el uso anteriormente descrito que los actores maliciosos hacen de los modelos de aprendizaje en el ciberespacio. En este apartado se desarrollarán de manera general otros diferentes usos de la IA en ciberdefensa, para tener una visión clara en el enfoque del proyecto.

- Detección temprana y análisis de amenazas

Uno de los campos en los que la IA ha demostrado tener un gran impacto es en la identificación de patrones anómalos en sistemas y redes militares, ayudando así a detectar comportamientos sospechosos, que de otra manera pasaría desapercibidos. Además, esta detección se realiza a velocidades que reducen drásticamente el tiempo de exposición a ataques, mejorando así la resiliencia de los sistemas. [25]

- Apoyo a la toma de decisiones

La capacidad de procesamiento de datos de la IA se aprovecha en el análisis de las campañas de ciberataques, con la correlación de incidentes dispersos en diferentes dominios, integrándolo todo en un cuadro de situación, que resulta muy útil en la coordinación multinivel de las operaciones militares complejas. [25]

- Lucha contra la desinformación y los deepfakes

Al igual que los ciberatacantes utilizan la IA para realizar estas campañas de desinformación, en ciberdefensa se utiliza la IA para detectar los contenidos generados por IA, con el desarrollo de algoritmos capaces de identificar patrones en imágenes, vídeos y textos. Estas técnicas permiten a los Estados responder con prontitud a intentos de injerencia en procesos importantes como elecciones o crisis. [26]

- Modelamiento y simulación en escenarios complejos

Con el objetivo de una instrucción y entrenamiento más avanzado en el ámbito del ciberespacio, se han desarrollados herramientas de modelamiento y simulación asistida por IA.

En el ámbito marítimo, se puede destacar el proyecto MARCIM (Marco de referencia para el modelamiento y simulación de la ciberdefensa marítima), impulsado por la Escuela Naval de Cadetes “Almirante Padilla” de Colombia. En este proyecto se han utilizados modelos matemáticos tales como la dinámica de sistemas o la teoría de juegos, para la representación de los atacantes y defensores. La simulación tiene como objetivo reproducir tanto la dimensión táctica y operacional como la estratégica, con la defensa de infraestructuras críticas. [17]

Más allá del sector marítimo, se han desarrollado iniciativas como Cyber-FIT[15], que desarrollan entornos multiagente donde las unidades cibernéticas se enfrentan en simulaciones que replican entornos de guerra cibernética realista. El uso y desarrollo de estos softwares permiten realizar predicciones más ajustadas de la evolución que podría tener un ataque de un grupo APT bajo múltiples condiciones.

- Automatización de la respuesta a incidentes

Los algoritmos de aprendizaje pueden llegar a seleccionar contramedidas adecuadas ante un incidente, sin requerir de intervención humana. Estas capacidades evitan ataques como ransomware o DDoS, que requieres reacciones en segundos para evitar que colapsen las infraestructuras.

4. DESARROLLO: ANÁLISIS Y RESULTADOS

4.1. Análisis de las necesidades en el ámbito de la Ciberdefensa

Con el objetivo de detectar las necesidades asociadas al ámbito de ciberdefensa en la integración de IA en plataformas SIEM, se ha elaborado una matriz DAFO, resumida en una tabla al final del capítulo, que sintetiza los principales aspectos que influyen en la adopción de estas necesidades. Por un lado, se analizan las debilidades y fortalezas que se asocian a factores internos de las integraciones SIEM-IA y por otro, las amenazas y oportunidades que responden a condiciones externas del entorno estratégico, tecnológico y normativo. Para la elaboración de este análisis se han tenido en cuenta los resultados obtenidos del *Cuestionario de valoración sobre integración de IA en plataformas SIEM realizado a especialistas en ciberdefensa* (ver Anexo II).

[15] Cyber Forces Interaction Terrain es un framework para simular el rendimiento de equipos de ciberseguridad, que permite modelas y analizar cómo los equipos cibernéticos interactúan entre sí y con el entorno. Fuente: https://kilthub.cmu.edu/articles/thesis/Cyber-Forces_Interactions_Terrain_An_agent-based_framework_for_simulating_cyber_team_performance/21385434?file=37956423

Debilidades:

1. Una amplia mayoría de SIEM de código abierto carecen de IA nativa, lo que obliga a emplear soluciones externas, mediante APIs o servidores intermedios. Las plataformas que usan modelos de IA específicas para análisis de alertas y vulnerabilidades son comerciales (los ya vistos Watson y Splunk, o Microsoft Sentinel), por lo que la IA utilizada en integraciones externas no es específica y estandarizada. Esto limita la explicabilidad, reduciendo la confianza en los resultados obtenidos.
2. Si la integración se realiza con herramientas que hacen uso de la nube, la información que manejen puede estar comprometida. En el caso de la información que se maneja en este ámbito de monitorización que hacen los SIEM, esta debe manejarse localmente, sin riesgo a que las vulnerabilidades encontradas o las alertas procesadas puedan ser filtradas.
3. Si la integración se realiza localmente, el consumo de recursos internos del sistema es mucho mayor, principalmente por el volumen de recursos computacionales que los modelos de lenguaje locales requieren para gestionar las grandes cantidades de datos que manejan. Además, esta posee menos flexibilidad estructural que las herramientas en la nube.
4. Dificultad de homogeneización en los despliegues locales o en la nube, debido a que algunos SIEM se integran mejor con IA en la nube, mientras que otros permiten solo despliegues locales.
5. Hoy en día, no existe un estándar universal que defina como deben ser las comunicaciones entre las plataformas SIEM y los modelos de aprendizaje automático. Cada fabricante de SIEM y de cada solución IA ofrece sus propias herramientas (APIs, pipelines, protocolos de autenticación). Esto genera heterogeneidad excesiva en los formatos de datos de, por ejemplo, los logs y las alertas (JSON, XML, CEF), lo que dificulta a la IA la normalización de los datos antes de su procesamiento.
6. Cuando se usan herramientas civiles (de código abierto o comerciales no específicamente diseñadas para uso militar), se debe de realizar una estandarización de los entornos donde estas son desplegadas. Esto conlleva una carga administrativa y técnica, debido a que se debe invertir recursos en documentación y mantenimiento para la creación de este marco estandarizado.

Fortalezas:

1. La integración de modelos de IA automatiza la priorización de alertas, filtrando falsos positivos y destacando las más críticas, reduciendo la carga de trabajo de los analistas.
2. El aprendizaje en tiempo real y la adaptación es una de las principales fortalezas. La IA tiene la capacidad de aprender patrones históricos y de reconocer nuevos ataques con amenazas que no estaban contempladas en las reglas estáticas, basándose en alertas o incidencias anteriores. Esto mejora la resiliencia que se necesita para luchar contra el exponencial crecimiento y aparición de actores y grupos sofisticados en el ciberespacio.
3. Esta integración tiene una escalabilidad alta, permite escalar desde pequeñas redes hasta infraestructuras críticas complejas sin necesidad de rediseñar la arquitectura. Además, en el caso de que la infraestructura este diseñada con recursos adecuados, la plataforma permite gestionar grandes volúmenes de datos sin que el rendimiento del sistema se degrade.

4. Si se adoptan integraciones abiertas (flexibles o de código abierto), se evita la dependencia de un único fabricante, reduciendo el riesgo de "vendor lock-in"[16] y facilitando en este caso la integración de nuevas tecnologías o de cambios en el proveedor sin comprometer la continuidad del proyecto.
5. Haciendo uso de herramientas de código abierto o de libre distribución, se eliminan los costes asociados a licencias comerciales.
6. La integración local reduce los riesgos de exposición de los datos, manteniendo un mayor control y soberanía sobre estos, lo que resulta esencial en el sector de defensa y en la seguridad nacional, evitando de esta manera la exposición de información sensible a servicios externos. Por otro lado, las integraciones en la nube reducen el volumen de recursos computacionales, favoreciendo así la eficacia y elasticidad del procesamiento de datos en la plataforma.
7. La IA permite la integración de múltiples fuentes de información procedentes de diferentes bases de datos de vulnerabilidades para correlacionar de manera más eficiente los incidentes, mejorando el contexto y ofreciendo a los analistas una visión más amplia del problema.

Amenazas:

1. Las regulaciones sobre protección de datos e IA pueden limitar la adopción de ciertos enfoques. La proliferación de algunas normativas en Europa limita el uso de determinados modelos o prácticas de tratamiento de información. Esto puede afectar con restricciones en la implementación de IA en entornos y redes de defensa cuando los datos son sensibles.
2. Existe una presión aumentada sobre los SOC debido al uso de IA por actores maliciosos. La automatización de las campañas de ataques y el avance en la evasión a los sistemas de detección, incrementan la complejidad y el volumen de amenazas que los SIEM deben enfrentar.
3. En casos recientes de ataques, se ha evidenciado un aumento de estos en infraestructuras críticas, como servicios nacionales informáticos o satélites.
4. El crecimiento exponencial de las tecnologías genera una obsolescencia tecnológica que genera la necesidad constante de actualización. Las plataformas tanto SIEM como de IA pueden quedar desactualizando frente a nuevas técnicas de intrusión y frente a nuevos requisitos de hardware y software, lo que obliga en una inversión continua en mejoras.
5. Una de las amenazas principales es la concentración de las principales tecnologías en pocos proveedores, lo que deriva en una dependencia estratégica que dentro de las plataformas mencionadas pueden estar condicionadas a restricciones geopolíticas, cambios en las políticas de licencias o incluso interrupciones en la cadena de suministro digital de los desarrolladores.

Oportunidades:

1. Se está produciendo un desarrollo muy marcado en los algoritmos basados tanto en machine learning como en deep learning, que se pueden aplicar tanto en detección y clasificación de amenazas como en predicción. La detección temprana y autónoma incluye la de vectores de ataque y evolución de amenazas. Por otra

[16] El vendor lock-in es el uso restringido de una tecnología o servicio desarrollada por un proveedor, que se asegura de crear dependencias por sus servicios. Fuente: https://www.arsys.es/blog/que-es-vendor-lock-in-que-tipos-existen-y-como-se-puede-evitar-esta-situacion

parte, la detección de incidentes futuros incluye también las amenazas desconocidas o incluso comportamientos encubiertos que no son visibles con reglas estáticas.

2. Los avances en tecnologías de integración como MCP permiten que las conexiones entre SIEM y IA sean más flexibles y modulares, facilitando la experimentación con diferentes modelos de distintos proveedores.
3. La UE y España han lanzado estrategias que fomentan la aplicación de IA a la seguridad y defensa, como la Cybersecurity Strategy for the Digital Decade [27] o el plan nacional España Digital 2026 [28]. Con esto se acelera la incorporación de los modelos de lenguaje al ámbito de defensa y se crea un marco político favorable.
4. La cooperación internacional en materia de ciberseguridad e IA avanza a pasos agigantados y desde la OTAN y la UE se impulsan múltiples intercambios de prácticas, inteligencia y modelos de referencia, reforzando las capacidades conjuntas defensivas antes amenazas globales en el ciberespacio.
5. Mayor disponibilidad dentro de los servicios en la nube con la introducción del Edge Computing, procesos que acercan la nube a donde los datos están siendo generados. Este crecimiento favorece la elasticidad para el entrenamiento y modelado de lenguajes complejos.

Tabla 3 Matriz DAFO resumen de la integración de IA en plataformas SIEM. Fuente: Elaboración Propia

ASPECTOS INTERNOS	
Debilidades	*Fortalezas*
-> Ausencia de IA nativa en la mayoría de los SIEM de código abierto. *-> Uso de IA externa no estandarizada ni específica para los SIEM.* *-> Riesgo de exposición de información en casos de integraciones en la nube.* *-> Elevado costa computacional y menos flexibilidad estructural en las integraciones locales.* *-> Dificultad en la coordinación entre SIEM e IA con despliegues locales y en la nube.* *-> Falta de un estándar de integración SIEM-IA.* *-> Necesidad de estandarización de entornos al hacer uso de herramientas civiles.*	*-> Automatización de la priorización de alertas y reducción de falsos positivos.* *-> Adaptación a nuevas amenazas y capacidad de aprendizaje.* *-> Alta escalabilidad.* *-> Evita dependencia de un único proveedor.* *-> Con el uso de herramientas de libre distribución se evita el coste de licencias.* *-> Mientras las integraciones locales aportan confidencialidad en los datos, la nube aporta flexibilidad.* *-> La capacidad de integrar diversas fuentes enriquece las respuestas.*
ASPECTOS EXTERNOS	
Amenazas	*Oportunidades*
-> Restricción del marco legal en materia de protección de datos e IA. *-> Uso de IA por ciberdelincuentes.* *-> Incremento de ciberataques en estructuras críticas.* *-> Necesidad de actualizaciones constantes en las tecnologías por el riesgo de obsolescencia.* *-> Dependencia estratégica de pocos proveedores.*	*-> Avance de los algoritmos de IA aplicables a este campo.* *-> Desarrollo de tecnologías de integración, haciendo estas más flexibles y modulares.* *-> Impulso de las estrategias y planes nacionales y de la UE en materia de seguridad y defensa.* *-> Fomento de las cooperaciones internacionales en IA y ciberseguridad.* *-> Mayor disponibilidad y crecimiento de los servicios en la nube.*

4.2. Justificación Técnica

Para el desarrollo práctico de este TFG, se ha llevado a cabo una selección de herramientas tecnológicas, ya que de estas depende tanto la eficacia de detección como la capacidad de respuesta. Se ha planteado importante realizar una comparativa estructurada de las diferentes soluciones IA y diferentes plataformas SIEM, considerando distintos criterios de coste, despliegue, rendimiento y en especial viabilidad de integración con plataformas SIEM en un entorno local. La información empleada en estas comparativas se ha obtenido de las fuentes oficiales de cada proveedor, a fin de garantizar la veracidad y actualidad de los datos. [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]

4.2.1. Comparativa de soluciones de IA existentes

- Análisis de los diferentes modelos

Llama 3 (Meta)

Ofrece versiones descargables bajo licencia propia de Meta (Llama 3 Community License), permitiendo la ejecución local tras aceptarla, con condiciones de uso como limitaciones de redistribución y casos de uso sensibles. [30] Permite la descarga y ejecución en local, siempre que se cumplan los requisitos de licencia. Su rendimiento en generación y razonamiento es muy bueno, sin embargo, sus modelos de mayor tamaño requieren una gran capacidad de cómputo, lo que puede limitar su aplicación en un entorno de laboratorio para fines académicos con hardware moderado.

Mistral

Mediante el software Ollama, que permite la ejecución de modelos de IA directamente en una máquina local, podemos descargar el modelo abierto de mistral, que cuenta con licencia Apache-2.0, apto para ejecución en local, sin salida de datos al exterior y garantía de seguridad. Ofrece buenas prestaciones y equilibrio entre rendimiento y eficiencia, pudiendo ejecutarse incluso en hardware moderado con equipos de gama media, lo que lo hace especialmente adecuado para proyectos de laboratorio académico. Al ser un modelo de código abierto, su coste de uso es gratuito. En cuanto a integración con otras plataformas, ofrece muy buenos resultados con APIs locales y otras herramientas similares.

Claude

Los distintos modelos de esta serie se encuentran entre los más avanzados en cuanto a razonamiento y generación de texto natural. Su uso está limitado sin embardo a través de API, únicamente con despliegue en la nube, lo que no permite la confidencialidad de los datos que maneje. La versión que soporta API es Claude Pro, su coste varía dependiendo de la cantidad de texto y datos que se procesen (pago por tokens), con diferentes planes dependiendo del volumen requerido. Existe otra posibilidad gratuita para la integración con otras herramientas, la implementación de ésta dentro de Docker Desktop y el uso de MCP, de lo cual se hablará en apartados posteriores. Estos modelos son referencia en materia de razonamiento complejo y son idóneos para tareas que requieren un elevado grado de comprensión textual.

ChatGPT / OpenAI

OpenAI ofrece sus modelos GPT exclusivamente a través de API en la nube o con Azure OpenAI Service. Esto imposibilita su despliegue local, ya que cualquier integración vía API enviaría sus datos a los servidores de OpenAI. Su coste al igual que Claude es de pago por uso, disponiendo de versión gratuita. Su precisión y versatilidad impone este modelo como uno de los más avanzados, siendo una de las mejores opciones si no fuera por el compromiso de información que supondría el egreso de datos y dependencia de un proveedor externo para el entorno de defensa.

Deepseek

Es un modelo con un alto rendimiento y coste muy reducido, ganando por ello popularidad y siendo una iniciativa emergente en la actualidad. Aunque existen intentos para su despliegue local, la posibilidad de la API sigue siendo egresando datos a la nube, esta API es comercial, con costes más bajos que Claude y ChatGPT, aunque carece de la madurez y el ecosistema que poseen Mistral y Llama.

- Criterios de evaluación

La adopción de estos modelos de lenguaje requiere un análisis de los factores críticos, para ellos se han establecido los siguientes criterios, con un peso porcentual, reflejando así su importancia en el contexto del TFG:

1. Despliegue local (20%): capacidad del modelo de ejecutarse en entornos propios sin depender de la nube, garantizando así la confidencialidad de la información.
2. Licencia y coste de uso (20%): en función de la gratuidad de los servicios y de la existencia de costes recurrentes como tokens, licencias o restricciones legales.
3. Rendimiento (20%): teniendo en cuenta la velocidad, la eficiencia en un hardware limitado y el consumo de recursos.
4. Calidad de generación (15%): precisión y coherencia de las respuestas en relación con la utilidad para un SIEM.
5. Capacidad de adaptación (15%): capacidad para adaptar el modelo a tareas específicas.
6. Ecosistema e integración (10%): disponibilidad de herramientas para realizar la integración de manera más simple.

Basándonos en el análisis anterior se ha elaborado una tabla puntuando los criterios de evaluación propuestos, para de esta forma poder seleccionar la herramienta más adecuada para este trabajo.

Tabla 4 Matriz de prestaciones de los modelos de IA. Fuente: Elaboración propia

Modelo	*Despliegue local*	*Licencia / coste*	*Rendimiento*	*Calidad*	*Adaptación*	*Integración*	*Total*
Mistral	*5*	*5*	*4*	*4*	*4*	*5*	*4,5*
Llama 3	*4*	*4*	*4*	*5*	*4*	*4*	*4,15*
OpenAI	*1*	*2*	*5*	*5*	*3*	*4*	*3,2*
Claude	*1*	*2*	*5*	*5*	*3*	*4*	*3,2*
DeepSeek	*2*	*3*	*4*	*4*	*3*	*4*	*3,25*

Para complementar la comparación objetiva de estas soluciones evaluadas también se va a aplicar la metodología AHP (Analytic Hierarchy Process), con el fin de enriquecer la comparativa y ponderar los criterios definidos en la matriz de prestaciones. El desarrollo del análisis AHP se incluye en el Anexo III. A continuación, se muestra la matriz de decisiones con los resultados obtenidos.

MATRIZ DE DECISIÓN

CRITERIOS / SUBCRITERIOS	PESOS	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Despliegue local	0,26	0,50	0,30	0,05	0,05	0,10
Licencia/coste	0,26	0,50	0,26	0,05	0,05	0,13
Rendimiento	0,26	0,11	0,11	0,33	0,33	0,11
Calidad	0,10	0,09	0,27	0,27	0,27	0,09
Adaptación	0,10	0,33	0,33	0,11	0,11	0,11
Integración	0,04	0,43	0,14	0,14	0,14	0,14
		0,34	**0,24**	**0,15**	**0,15**	**0,11**

Figura 11 Matriz de decisión AHP de soluciones IA

- Elección

  Después de la evaluación comparativa de las diferentes opciones, la opción más adecuada para este entorno académico es Mistral, ejecutado mediante el software Ollama. Esta decisión se ha tomado prestando especial atención a su ausencia de costes y a la capacidad de su despliegue local, garantizando así a la soberanía de los datos.

  No obstante, con el objetivo de enriquecer el análisis y contraste de resultados, aportando diferentes escenarios tecnológicos, también se usará Claude, en su versión gratuita, para evaluar la arquitectura de conexión a través de Docker Desktop y MCP, aunque se trate de un servicio en la nube, con vistas a facilitar una futura adaptación en despliegues locales.

4.2.2. Comparativa de plataformas SIEM

- Análisis de las diferentes plataformas

Los SIEM están enmarcados dentro de un ecosistema llamado Monitorización de la Seguridad de la Red (NSM). Esta metodología clarifica que un SIEM no funciona de manera aislada, sino que se nutre de diferentes herramientas, permitiendo así el flujo de monitorización. Estas herramientas incluyen: captura y análisis de tráfico, sistemas de almacenamiento e indexación, plataformas de visualización y gestores de incidentes.

Las plataformas aquí analizadas proporcionan una estrategia distinta en cuanto a herramientas NSM y arquitectura de despliegue.

Security Onion

Es una distribución de Linux, de coste gratuito, especializada en NSM. Este SIEM se centra principalmente en la captura y análisis de tráfico de red. Su arquitectura se basa herramientas NSM muy reconocidas, como Suricata como motor IDS/IPS, Zeek para el análisis de tráfico y ElasticSearch y Kibana para indexación y visualización respectivamente. Este conjunto de componentes deriva en un SIEM muy potente y sólido. A nivel de integración con IA local, es posible aprovechar ElasticSearch para exportar alertas, además, su rendimiento es bueno en entornos controlados, sin embargo, para realizar la integración por medio de servicios de MCP, se necesita disponer de la versión de pago.

Splunk Enterprise Security

Es una de las soluciones comerciales más potentes, tiene capacidad para desplegarse tanto en entornos locales como en la nube. Aunque flexible, esta opción tiene un coste de licencia alto, una de sus principales desventajas. Su cobertura es muy completa, incluyendo analítica avanzada, UEBA y un motor de búsqueda y correlación que es una de sus principales fortalezas. Este motor está basado en lenguaje SPL (Search Processing Language), permitiéndole analizar en tiempo real enormes cantidades de datos. La recogida de información se realiza con sus propios endpoints y conectores y el almacenamiento se gestiona en índices propios optimizados.

La integración con IA es factible mediante APIs, con licencia limitada y costes. Su escalabilidad y rendimiento son muy sobresalientes, con capacidad de gestión de enormes volúmenes de datos. Además, es una referencia para los SOC de grandes empresas.

Wazuh

Se trata de un SIEM de código abierto, con la posibilidad de desplegar completamente en local y offline. Gracias a que no requiere licencias comerciales ni costes, es una opción totalmente gratuita y accesible. Dispone de agentes para endpoints y servidores, los datos recopilados mediante estos en materia de logs, inventario de activos, cambios en archivos e intentos de intrusión son posteriormente procesados por el manager tras ser indexados. Para el almacenamiento se usa OpenSearch, optimizada y con buena capacidad y para la visualización su propio Wazuh Dashboard, Su escalabilidad y rendimiento, aunque no son los mejores, resultan suficientes para entornos académicos de laboratorio, además se integran fácilmente con IA locales, tanto con APIs como directamente desde el indexador.

IBM QRadar

Este SIEM dispone, a diferencia de otras plataformas de un NSM nativo, llamado Flow Collector, que no depende de IDS/IPS externos, siendo por tanto independiente de herramientas de terceros. Esta plataforma además cuenta con distintos paquetes preconfigurados, haciéndolo atractivo para sectores que requieren una regulación y cumplimiento normativo, como en el caso de defensa. Otro punto a favor es la incorporación de una IA nativa, pero esta es en la nube y con coste. La integración no está concebida nativamente, lo que podría dificultar la tarea. El respaldo de la empresa IBM le aporta solidez y un soporte de alto nivel, aunque el coste de licencia lo convierte en una opción no tan correcta para este trabajo.

Microsoft Sentinel

Al ser un servicio de Microsoft Azure, no admite despliegue local. En lugar de licencias fijas, Sentinel cobra por GB de datos ingeridos, lo que es bueno para su escalabilidad y flexibilidad, pero los costes van a ser variables y difíciles de predecir. La integración con aplicaciones de otros ámbitos, como modelos de IA, es muy llevadera con herramientas que se encuentren dentro del ecosistema de Microsoft, sin embargo, por esta razón la integración con IA local es muy limitada. Su cobertura funcional es amplia, con capacidades UEBA integradas. Cuenta con una de las mayores redes de soporte y documentación.

- Criterios de evaluación

El análisis de los factores críticos de estas plataformas se ha fijado basándose en los siguientes factores críticos, con un peso porcentual, en consonancia con el contexto de este trabajo:

1. Despliegue local (20%): al igual que en los modelos de IA, capacidad del modelo para mantener un sistema offline y asegurando la soberanía de datos.
2. Licencia y coste de uso (20%): basada en la gratuidad y los posibles costes asociados.
3. Cobertura funcional (20%): posibles capacidades (detección, prevención, correlación, etc.)
4. Integración con IA local (15%): capacidad de exportar datos a un modelo de IA.
5. Escalabilidad y rendimiento (15%): manejar grandes volúmenes de datos sin perder eficacia.
6. Comunidad y soporte (10%): respaldo de la comunidad y documentación disponible.

De igual forma que para los modelos IA, basándose en el análisis se proponen las siguientes puntuaciones para los criterios de evaluación, expuestos en la siguiente tabla:

Tabla 5 Matriz de prestaciones de las plataformas SIEM. Fuente: Elaboración propia

Plataforma	*Despliegue local*	*Licencia / coste*	*Cobertura*	*IA local*	*Escalabili-dad*	*Comunidad*	*Total*
Wazuh	*5*	*5*	*3*	*5*	*3*	*4*	*4,17*
Security Onion	*5*	*4*	*3*	*4*	*4*	*4*	*4*
Splunk	*1*	*1*	*5*	*4*	*5*	*5*	*3,5*
IBM QRadar	*1*	*1*	*5*	*3*	*4*	*5*	*3,17*
Sentinel	*1*	*1*	*5*	*1*	*5*	*5*	*3*

Siguiendo la misma línea que la selección de soluciones IA, se ha realizado un análisis AHP, con el fin de ponderar cuantitativamente los factores definidos en la matriz de prestaciones. El desarrollo del análisis AHP se incluye en el Anexo IV. A continuación, se muestra la matriz de decisiones con los resultados obtenidos.

MATRIZ DE DECISIÓN

CRITERIOS / SUBCRITERIOS	PESOS	Wazuh	Security	Splunk	IBM QRadar	Sentinel
Despliegue Local	0,26	0,41	0,41	0,06	0,06	0,06
Licencia/coste	0,26	0,50	0,33	0,06	0,06	0,06
Cobertura	0,26	0,06	0,06	0,29	0,29	0,29
IA local	0,10	0,45	0,21	0,21	0,10	0,04
Escalabilidad	0,10	0,06	0,13	0,34	0,13	0,34
Comunidad	0,04	0,09	0,09	0,27	0,27	0,27
		0,30	**0,24**	**0,17**	**0,14**	**0,15**

Figura 12 Matriz de decisión AHP de soluciones SIEM

- Elección

Después de realizar el análisis y la matriz de prestaciones, se observa que las plataformas comerciales proporcionan una cobertura funcional muy amplia y con un excelente rendimiento, sin embargo, estos factores no compensan su coste y dependencia de la nube. En este marco, Wazuh se posiciona como la mejor elección, debido al equilibrio que posee entre rendimiento y coherencia con los objetivos del proyecto, aunque sus capacidades no alcancen el nivel de las soluciones comerciales más avanzadas.

4.3. Diseño

4.3.1. Componentes de Wazuh

Con el objetivo de comprender la arquitectura de la integración propuesta, es importante explicar cómo es el ecosistema Wazuh y cuáles son las partes de las que se compone.

Componentes principales:

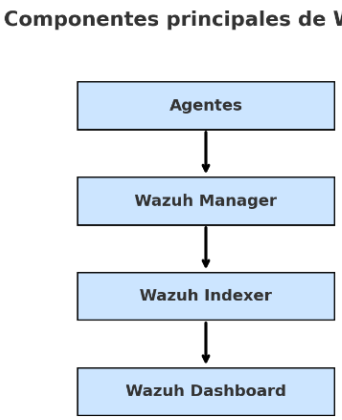


Figura 13 Componentes principales de Wazuh. Fuente: ChatGPT

- Agentes Wazuh: Son programas ligeros instalados en dispositivos finales para monitorizarlos, se encargan de recoger datos de seguridad como logs del sistema, integridad de archivos o actividades del usuario. Estos datos son la materia prima que será utilizada más tarde para ser procesada y correlacionada y generar incidencias.
- Wazuh Manager: Es el núcleo del sistema y se encarga de recibir, procesar y correlacionar los eventos provenientes de los agentes o del propio equipo donde esté instalado con la ayuda de un motor propio de correlación de reglas. Además, también administra las políticas de seguridad y gestiona los agentes.
- Wazuh Indexer: Basado en OpenSearch/Elasticsearch, se encarga de almacenar e indexar los eventos procesados por el Manager. Elementos clave para una integración externa debido a su capacidad de extracción de alertas.
- Wazuh Dashboard: Es la herramienta que proporciona la interfaz de visualización, construida sobre Kibana, permite ver gráficamente los eventos, alertas, reglas y métricas. Permite dentro del contexto de este trabajo, validar los datos enviados a los modelos.

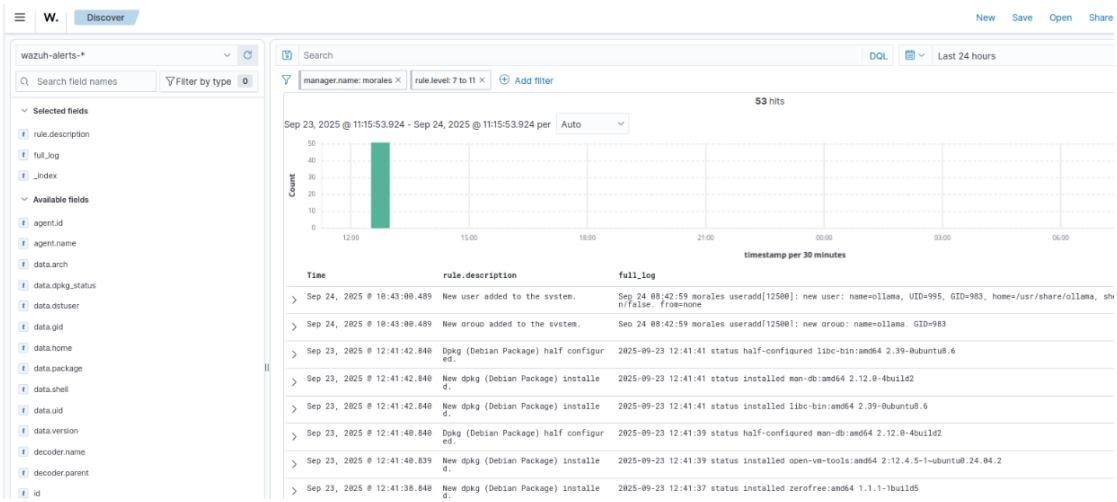


Figura 14 Ejemplo de la interfaz de visualización de alertas en el dashboard de Wazuh. Fuente: Elaboración propia

4.3.2. Desarrollo de los modelos propuestos

El diseño preliminar del sistema se centra en definir las diferentes opciones de integración entre la plataforma de Wazuh y modelos de IA tanto locales (Mistral) como externos (Claude). Para ello se plantean tres aproximaciones, con el objetivo de establecer un marco para realizar distintas evaluaciones de viabilidad en el siguiente apartado. Estas aproximaciones son: un modelo teórico basado en la integración mediante API REST en un entorno Linux; un modelo práctico aplicado, que soluciona las limitaciones del modelo anterior, usando en este caso el indexador de Wazuh; y un modelo experimental, en este caso desplegado en un entorno de Windows, utilizando arquitecturas más novedosas con MCP Server como capa intermedia de comunicación entre el SIEM y la IA, este modelo contempla tanto la conexión con clientes externos como la securización de un modelo local. Las fuentes principales para el desarrollo del trabajo son las documentaciones oficiales de Wazuh [34], Ollama (Mistral) [29] y Centro Criptológico Nacional [39], que constituyen la base técnica y normativa de referencia.

Para los primeros modelos se usa un sistema operativo Ubuntu 24.04, con Wazuh 4.13 y Ollama (Mistral) desplegados como SIEM y motor de lenguaje, respectivamente, ambos instalados en local dentro del entorno. En el desarrollo se utilizó la terminal de Ubuntu (CLI, Command Line Interface), desde la cual se ejecutan los comandos necesarios, además se utilizan scripts de Python para automatizar las interacciones con las herramientas.

- Modelo teórico inicial: Integración vía API Manager en Ubuntu (fallido)

En una primera aproximación se planteó la posibilidad de obtener las alertas mediante la API REST de Wazuh Manager.

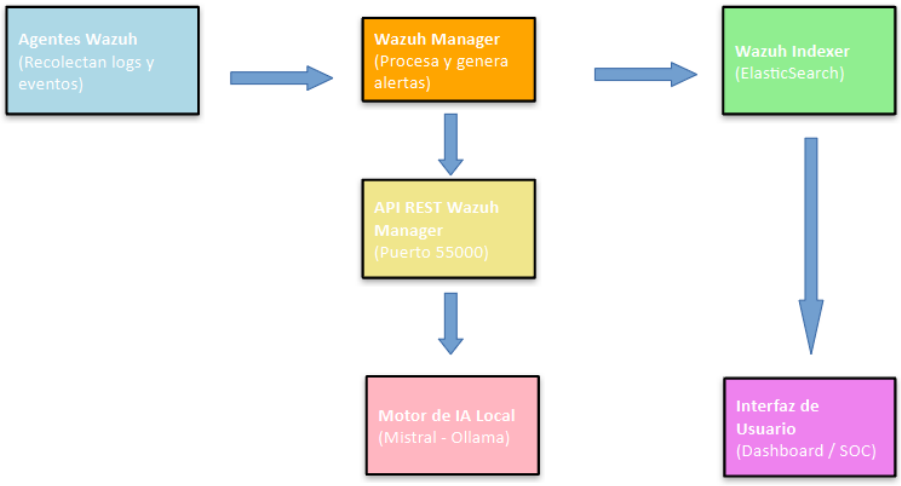


Figura 15 Esquema de la Arquitectura para el Sistema Propuesto. Fuente: Elaboración Propia

Aquí se presenta el diseño inicial para establecer la integración y permitir el flujo de comunicación entre Wazuh y el modelo de IA Mistral.

Los datos recolectados y procesados por el Wazuh Manager tiene dos vertientes, una es la indexación de estos datos por el Wazuh Indexer a la Interfaz de Usuario (Wazuh Dashboard) y otra es la capa de intermediación de la API REST de Wazuh Manager, que a su vez está conectada al modelo de IA local Mistral.

Desarrollo del proceso:

1. Comprobación del Dashboard: Se realizó una comprobación de herramienta para visualizar si las alertas e incidencias estaban siendo generadas por el Manager e indexadas por el Indexer. Para ello se accedió a través de un navegador web a la dirección “https://localhost”. Un ejemplo del contenido está en la Figura 12.

2. Comprobación de la API: Esta herramienta en Wazuh está expuesta en el puerto 55000 y es el punto de acceso donde se encuentran los datos gestionados por el Wazuh Manager. A través de la terminal de Ubuntu se envía una petición a la API, para obtener un token JWT (JSON[17] Web Token).

   Ejemplo de petición:

curl -u usuario:contraseña -k -X POST "https://localhost:55000/security/user/authenticate"

   “curl”: se usa para transferir datos desde o hacia un servidor.

   “-u”: para especificar que se va a usar usuario y contraseña.

   “-k”: se usa para permitir los certificados usados por la API de Wazuh

   “-X POST”: indica que se emplea el método HTTP POST, necesario para enviar las credenciales al servidor y obtener el token JWT de autenticación.

   “https://localhost:55000/security/user/authenticate”: es la dirección de autenticación de la API de Wazuh Manager expuesta en el puerto 55000 del propio equipo (localhost).

   Respuesta obtenida:

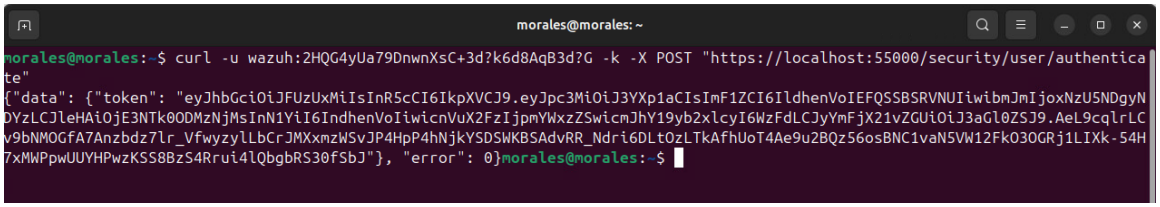


Figura 16 Respuesta obtenida de la autenticación de la API Wazuh Manager. Fuente: Elaboración Propia

   Con esta verificación se demuestra que la API responde de manera adecuada y que el token JWT generado permite autenticarse de forma válida.

3. Acceso a las alertas: Para consultar las alertas se probaron las rutas:

   https://localhost:55000/alerts

   https://localhost:55000/archives

   https://localhost:55000/security/events

   Todas ellas correspondientes a la API REST de Wazuh Manager, sin embargo, dichas peticiones resultaron fallidas. Tras realizar una investigación y revisión detallada de la documentación oficial, se constató que en la API del Manager de las nuevas versiones de Wazuh, no se ofrece acceso directo a las alertas almacenadas, sino únicamente a información de gestión (usuarios, agentes, configuración y estado del sistema). Tras esto se invalidó el modelo teórico inicial, puesto que no era capaz de establecer una conexión entre las alertas y el modelo de IA.

[17] El JavaScript Object Notation es un formato de texto para la comunicación entre servidores y navegadores web, legible tanto por humanos como por máquinas.

- Modelo práctico aplicado (Integración vía Indexador en Ubuntu)

A raíz del intento fallido anterior, se planteó un segundo enfoque que corrigió el problema: utilizad como punto de acceso el Indexador de Wazuh (Elasticsearch) para obtener los eventos almacenados. Este enfoque se desplegó en el mismo entorno que el anterior modelo.

El proceso se basó en el uso de la API REST nativa del Indexador, expuesta en el puerto 9200. A través de este puerto se consultan los índices donde Wazuh almacena las alertas (wazuh-alerts-*). A diferencia de la API REST del Wazuh Manager, la autenticación se realiza por defecto mediante credenciales de Basic Auth (usuario y contraseña) en lugar de con tokens JWT.

A través de la siguiente tabla se busca facilitar la comprensión de las diferencias entre ambas APIs.

Tabla 6 Tabla comparativa entre la API de Wazuh Manager y la API de Wazuh Indexer

Característica	*API Wazuh Manager*	*API Wazuh Indexer*
Puerto	*55000*	*9200*
Tipos de datos expuestos	*Información de gestión y configuración (estados del sistema, reglas, activos)*	*Alertas y eventos indexados*
Autenticación	*Por defecto JWT (JSON Web Token)*	*Por defecto Basic Auth (Usuario y Contraseña)*
Uso principal	*Administración de Wazuh*	*Consulta y análisis de eventos almacenados*
Dependencia	*Forma parte del Wazuh Manager*	*Es una API de un servicio independiente (Elasticsearch) integrada en Wazuh Indexer*

Esta será la nueva arquitectura. Esta vez es el Wazuh Indexer el que tiene dos vertientes, una para la Interfaz de Usuario (Dashboard) y otra para que la misma información pueda ser accedida a través de su API para llevar al motor local de Mistral.

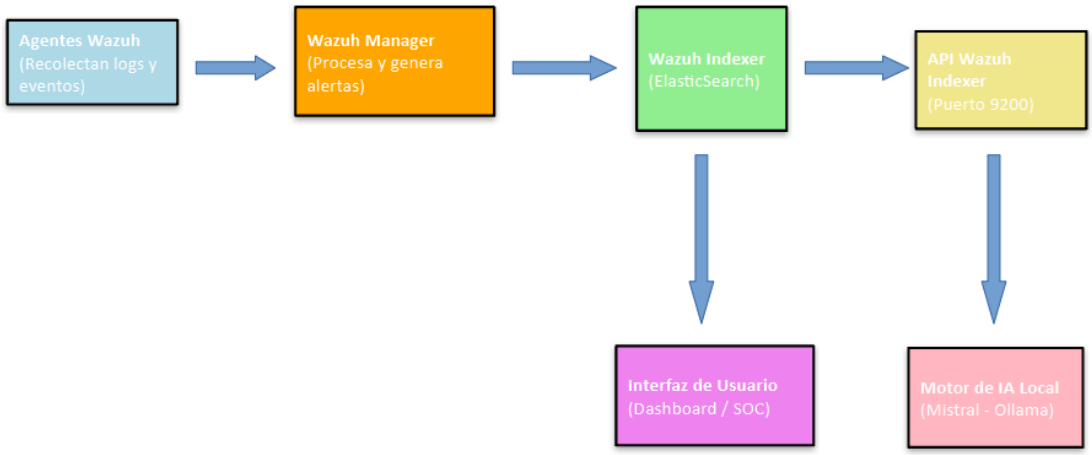


Figura 17 Esquema de la nueva Arquitectura para el Sistema Propuesto. Fuente: Elaboración Propia

Desarrollo del proceso:

1. Comprobación del Dashboard: Se verifica de igual manera que en el modelo anterior, para comprobar que las alertas están siendo generadas e indexadas.
2. Comprobación de la API del Indexer: Esta vez la herramienta está expuesta en el puerto 9200. De la misma manera que en el modelo anterior, se realiza una autenticación a través de la terminal de Ubuntu.

   Ejemplo de petición:

   curl -u usuario:contraseña -k -X GET “https://localhost:9200”

   “-X GET”: especifica el método HTTP GET, utilizado para realizar peticiones de lectura o consulta

   “https://localhost:9200”: dirección del servicio de indexación del Wazuh Indexer, alojado en el propio equipo (localhost)

   Respuesta obtenida:

Figura 18 Respuesta obtenida de la API Wazuh Indexer. Fuente: Elaboración Propia

   Esta vez nos devuelve una respuesta en formato JSON, que confirma el correcto funcionamiento de la API de Wazuh Indexer.

3. Obtención de alertas: Mediante un comando manual en la terminal de Ubuntu se envía una consulta a la API del Indexador.

   Ejemplo de consulta:

   curl -u admin:contraseña -k -X GET "https://localhost:9200/wazuh-alerts-*/_search?size=1&pretty"

   “-X GET”: se usa para especificar el método HTTP usado, en este caso GET.

   “https://localhost:9200/”: es la dirección de la API de Wazuh Indexer expuesta en el puerto 9200 del propio equipo (localhost).

   “wazuh-alerts-*/_search?size=1&pretty": se indica para la búsqueda (search) en el índice de alertas de wazuh, (wazuh-alerts-*), con los parámetros size=1 para que muestre solo una alerta y ?pretty para que la respuesta JSON sea con indentación y saltos de línea, facilitando la lectura por humanos.

Respuesta obtenida:

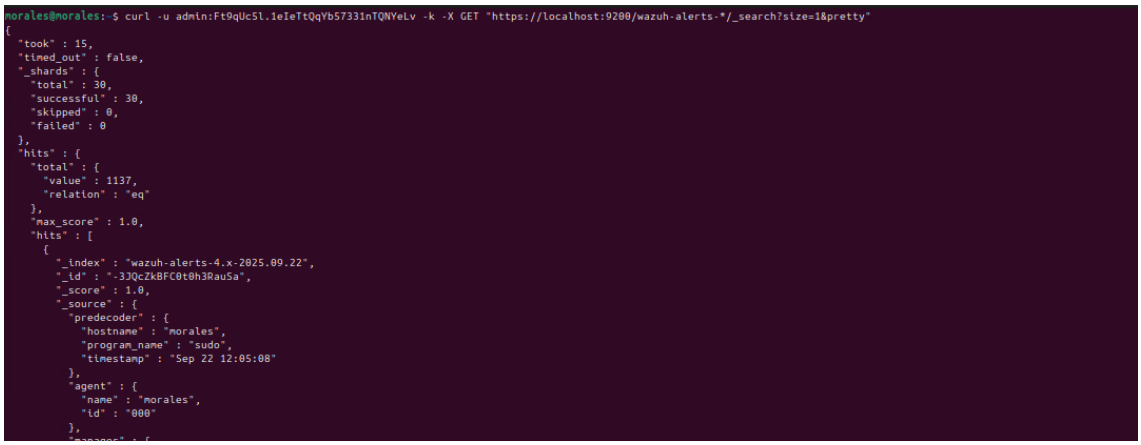


Figura 19 Obtención de alerta de la API Wazuh Indexer. Fuente: Elaboración Propia

4. Comprobación de la API de Ollama: Antes de enviar las alertas al modelo, se comprobó que la API de Ollama estuviese correctamente desplegada y accesible. Al usar Ollama solo en el entorno local a red, su API expuesta en el puerto 11434, no requiere autenticación.

   Ejemplo de petición:

   curl http://localhost:11434/api/tags

   Respuesta obtenida:

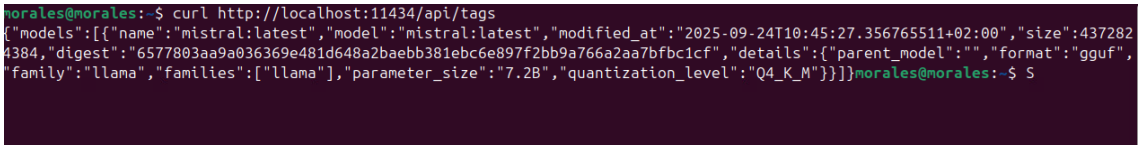


Figura 20 Respuesta obtenida de la API de Ollama. Fuente: Elaboración Propia

5. Interacción con Mistral: Con la obtención anterior de las alertas, se probaron comandos simples para enviarlas a Mistral y comprobar la viabilidad de la integración.

   Ejemplo de consulta:

   curl http://localhost:11434/api/generate -d '{

   "model": "mistral",

   "prompt": "Resume esta alerta: {contenido alerta}"

   }'

   La respuesta del modelo contiene un resumen textual de la alerta y sirvió para comprobar la viabilidad de la integración.

- Modelo experimental (Integración vía MCP Server en Windows)

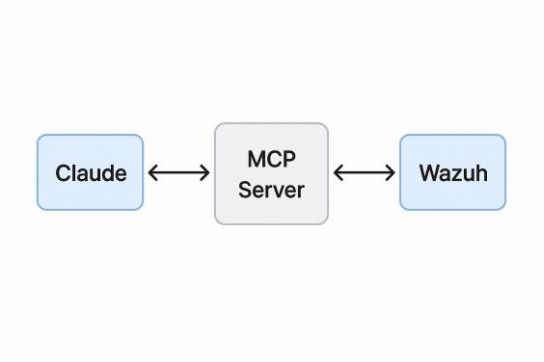


Figura 21 Arquitectura submodelo experimental A. Fuente: ChatGPT

Este modelo se diseñó con el objetivo de probar una arquitectura de integración entre el SIEM de Wazuh y un modelo de IA externo (Claude) mediante un servidor intermedio MCP. La idea principal expuesta en la arquitectura de la Figura 21, consiste en establecer un flujo bidireccional, en el que el MCP Server actúa como capa intermedia. Se utilizó la IA Claude, dado que el protocolo MCP es compatible con ella y porque Docker Desktop incorpora soporte nativo para la conexión mediante este protocolo.

El despliegue de este entorno se realizó en Windows 11 utilizando Docker Desktop, una herramienta que permite ejecutar contenedores Linux dentro de sistemas Windows o macOS mediante virtualización ligera. Estos contenedores funcionan como servicios aislados que comparten recursos del sistema anfitrión.

Desarrollo del proceso:

1. Preparación del entorno: se prepararon un grupo de contenedores pertenecientes a la suite de Wazuh (Manager, Indexer y Dashboard), cada uno en su contenedor correspondiente. Además, se añadió el contenedor adicional del servidor MCP de Wazuh, descargado desde el repositorio oficial del proyecto. [40] Esta estructura modular permite que cada componente funcione de manera aislada, pero comunicándose entre sí a través de la red interna de Docker.

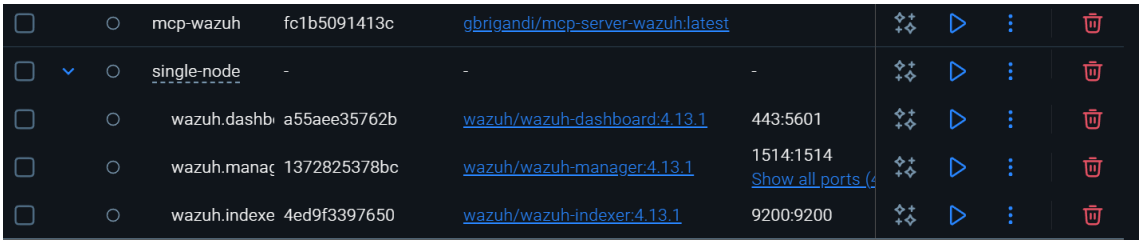


Figura 22 Contenedores instalados en Docker Desktop. Fuente: Elaboración propia

2. Configuración del MCP Server: como se ha señalado anteriormente, el despliegue de este servidor se ha realizado en un contenedor Docker. La configuración del servidor se ha llevado a cabo conforme a la guía del repositorio oficial del proyecto. [40] Ésta se gestiona a través de variables de entorno definidas en un archivo .env. que contiene los parámetros necesarios para conectarse a Wazuh Manager y Wazuh Indexer: usuarios, contraseñas, direcciones IP y puertos de conexión.

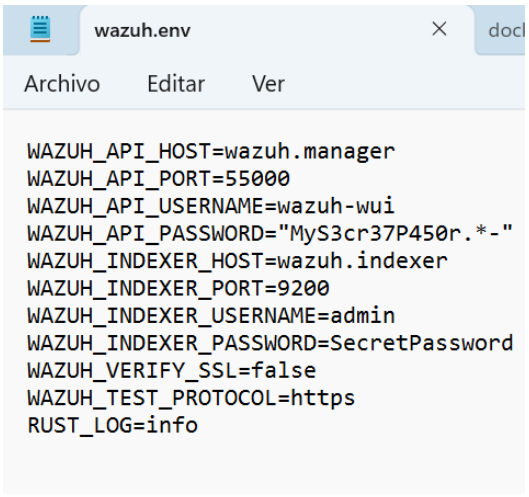


Figura 23 Archivo .env. Fuente: Elaboración propia

Estas variables no son leídas directamente por el servidor, sino que son interpretadas por Claude Desktop, utilizándolas para ejecutar el contenedor del MCP Server cuando Claude se ejecuta. De esta manera MCP accede a las credenciales y rutas necesarias para conectarse a Wazuh.

3. Conexión con Claude Desktop: para realizar la conexión de Claude con la infraestructura, fue necesario editar el archivo de configuración claude_desktop_config.json. En este archivo se especifica la ruta del archivo .env y los argumentos que permiten que, al iniciar Claude, se levante automáticamente el contenedor del MCP Server.

   El flujo quedaría de la siguiente manera:

   1. El usuario inicia Claude Desktop
   2. Claude ejecuta el contenedor Docker del MCP Server y le pasa los parámetros del archivo .env.
   3. El MCP Server se conecta al entorno de Wazuh.
   4. Claude establece comunicación con el MCP y obtiene la información del SIEM a través de él.

4.3.3. Pruebas de los modelos y análisis de resultados

- Modelo práctico aplicado (Integración vía Indexador en Ubuntu)

Para automatizar el flujo completo, desde la extracción de alertas hasta su clasificación y resumen, se desarrolló con la ayuda de ChatGPT un script de Python analizador_INDEXADOR.py

```python
import requests
import json

# --- Configuración ---
ES_URL     = "https://localhost:9200/wazuh-alerts-*/_search"
ES_USER    = "admin"
ES_PASS    = "Ft9qUc5l.1eIeTtQgYb57331nTQNYeLv"
N_ALERTS   = 5
OLLAMA_URL = "http://localhost:11434/api/generate"
MODEL      = "mistral"

# Desactivar warnings SSL (certificado autofirmado)
requests.packages.urllib3.disable_warnings()

# 1) Consulta al indexador (últimas N alertas ordenadas por timestamp descendente)
query = {
    "size": N_ALERTS,
    "sort": [{"@timestamp": {"order": "desc"}}]
}

resp = requests.get(
    ES_URL,
    auth=(ES_USER, ES_PASS),
    headers={"Content-Type": "application/json"},
    data=json.dumps(query),
    verify=False
)

if resp.status_code != 200:
    print("[ERROR] No se pudieron obtener alertas del indexador:", resp.text)
    exit(1)

hits = resp.json()["hits"]["hits"]

# Extraer solo el _source de cada alerta
alerts = [h["_source"] for h in hits]

print(f"Se han obtenido {len(alerts)} alertas del indexador")

# 2) Construir prompt para Mistral
prompt = (
    "Eres un analista SOC. Te paso alertas de Wazuh en JSON. Para que me contestes en español"
    "Clasificalas por severidad usando rule.level (0-3=bajo, 4-7=medio, 8-10=alto, 11+=critico)."
    "Asocia las alertas a un CVE"
    "Devuélveme un resumen con:\n"
    "- Total por severidad\n"
    "- 2-3 observaciones importantes\n\n"
    f"ALERTAS:\n{json.dumps(alerts, indent=2, ensure_ascii=False)}"
)

# 3) Enviar a Ollama (Mistral)
ollama = requests.post(
    OLLAMA_URL,
    json={"model": MODEL, "prompt": prompt, "stream": False}
)

if ollama.status_code != 200:
    print("[ERROR] Falló la llamada a Ollama:", ollama.text)
    exit(1)

print("\n--- Respuesta de Mistral ---\n")
print(ollama.json()["response"])
```

Figura 24 Script analizador_INDEXADOR.py. Fuente: Elaboración propia con ayuda de ChatGPT (OpenAI)

Realiza de manera automática la consulta al indexador de las alertas, la construcción del prompt donde incluye las alertas en formato JSON, el envío del prompt a Mistral y la recepción y visualización del resultado.

Respuesta obtenida:

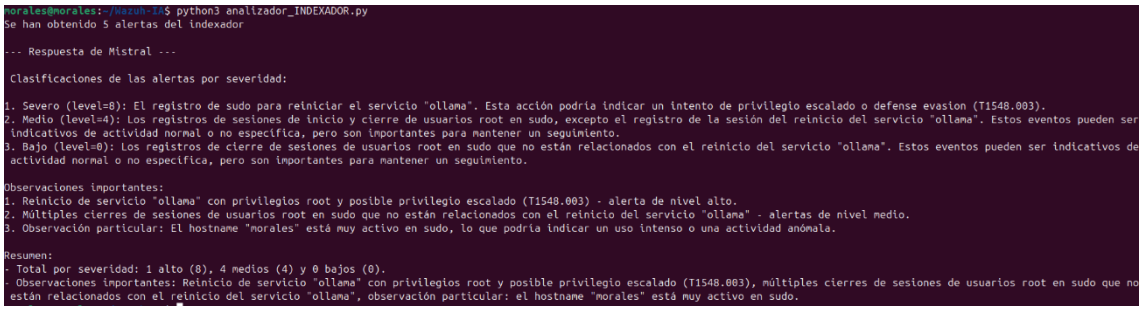


Figura 25 Respuesta obtenida de la ejecución del Script analizador_INDEXADOR.py. Fuente: Elaboración propia

Este paso constituye la culminación del modelo práctico, logrando un proceso automatizado de análisis SOC con IA, totalmente en local y sin dependencia de servicios en la nube.

- Modelo experimental (Integración vía MCP Server en Windows)

  Durante las pruebas de conexión entre Claude Desktop y el contenedor del MCP Server, se observó un error que, de manera recurrente, afectó a la conexión entre ambos, tal y como se muestra en la Figura 24. Tras analizar el registro del proceso y la configuración de ambos entornos, se determinó que la causa más probable se encuentra en una incompatibilidad entre los mecanismos de autenticación empleados por cada componente: Claude usa JSON Web Signature, mientras que el MCP Server de Wazuh, implementa autenticación mediante Basic Auth. Al intentar validar las credenciales transmitidas a través del archivo .env, el intercambio de tokens no se reconoce correctamente. Se considera por tanto una limitación técnica ajena al control del entorno experimental.

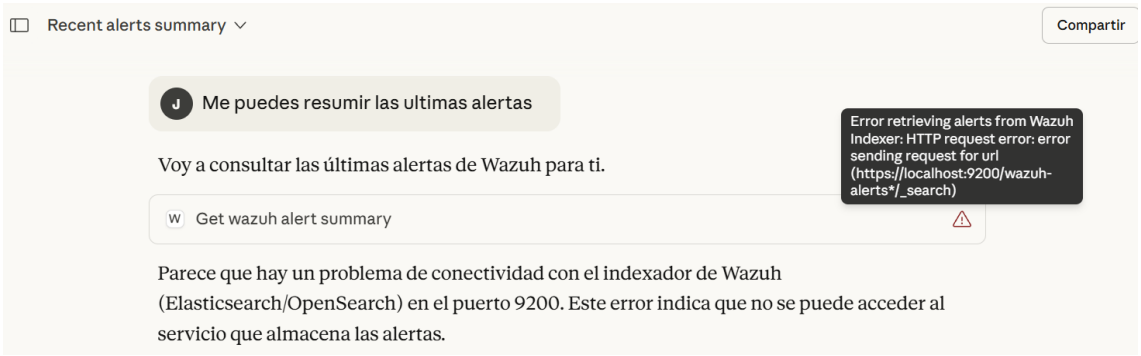


Figura 26 Error ejecución del contenedor MCP Server. Fuente: Elaboración propia

4.4. Viabilidad

4.4.1. Evaluación de viabilidad

Se ha realizado un estudio de viabilidad basado en el modelo TELOS. [41] Este análisis se ha llevado a cabo con el objetivo de estudiar la factibilidad del modelo práctico aplicado (integración de Wazuh vía API con Mistral) al que se llamará modelo A y el modelo experimental (Wazuh integrado mediante MCP Server a Claude) al que se llamará modelo B. Con este método se consigue evaluar de manera estructurada los factores más importantes que pueden ser condicionantes en la implementación de un proyecto tecnológico.

A continuación, se muestra el análisis TELOS, en el que se evalúan los puntos fuertes, limitaciones y riesgos de los factores tanto técnicos y económicos como legales, operativos y de planificación, representado en una tabla por cada dimensión.

Tabla 7 Análisis TELOS: Dimensión técnica. Fuente: Elaboración propia

DIMENSIÓN TÉCNICA		
Factor evaluado	**Modelo A**	**Modelo B**
Compatibilidad tecnológica	Alta – Integración directa mediante API Rest y Python, sin dependencias externas.	Media – Requiere MCP como capa intermedia y Docker
Recursos Hardware	Media – Uso intensivo de CPU por la IA local.	Alta – La IA externa conlleva una carga distribuida, menor consumo local
Escalabilidad y mantenimiento	Alta – Fácil de replicar	Media – Su arquitectura Docker es muy modular pero su escalabilidad es limitada
Soporte técnico y comunidad	Media-Alta – Amplia comunidad de ambas herramientas al ser de código abierto	Media – Existe una documentación limitada del MCP Server
Modularidad y madurez del desarrollo	Baja– Estructura primitiva en script, sin separación clara de servicios y sin interfaz	Alta – Arquitectura basada en contenedores Docker, facilitan el despliegue y actualización independiente

Tabla 8 Análisis TELOS: Dimensión económica. Fuente: Elaboración propia

DIMENSIÓN ECONÓMICA		
Factor evaluado	**Modelo A**	**Modelo B**
Coste de implementación	Medio-Bajo coste – Debido al uso de software libre. Sin embargo, la escalada de capacidad de la IA podría implicar gastos mayores	Medio-Alto coste – Potencial coste de IA debido a los recursos en la nube
Coste de mantenimiento	Bajo coste – Debido a la gran comunidad abierta de soporte	Medio coste – La conectividad con los servicios externos y la compatibilidad entre versiones puede incrementar el esfuerzo de mantenimiento
Financiación y recursos	Viabilidad Alta – Viable con presupuesto muy reducido. No requiere inversión ni pagos por uso una vez desplegado	Viabilidad Media – Requiere financiación sostenida a medio plazo debido a la disponibilidad de los recursos externos (Acceso a internet, coste de IA)

Tabla 9 Análisis TELOS: Dimensión legal. Fuente: Elaboración propia

DIMENSIÓN LEGAL		
Factor evaluado	**Modelo A**	**Modelo B**
Protección de datos	Alta – Datos locales sin transferencia a terceros	Media-Baja – Posible tratamiento de datos fuera del entorno
Licenciamiento software	Alto – 100% código abierto	Medio – Claude bajo licencia cerrada
Cumplimiento normativo potencial	Alto – Posible implementación de directrices CCN-STIC para la seguridad en entornos virtualizados y para la securización de la IA	Bajo – Ni Claude ni el servidor usado de MCP dispone de certificaciones CCN

Tabla 10 Análisis TELOS: Dimensión operacional. Fuente: Elaboración propia

DIMENSIÓN OPERACIONAL		
Factor evaluado	**Modelo A**	**Modelo B**
Facilidad de implementación	Alta – Despliegue rápido y reproducible	Media – Más complejo con múltiples servicios
Curva de aprendizaje	Media – Conocimientos básicos de Python y API REST	Media – Requiere conocimientos en Docker y MCP
Integración con SOC / SIEM	Alta – Conexión directa	Alta – Conexión directa
Soporte de operación 24/7	Alta – Control total del entorno al ser local	Media-Alta – Supervisión de contenedores

Tabla 11 Análisis TELOS: Dimensión planificación. Fuente: Elaboración propia

DIMENSIÓN PLANIFICACIÓN		
Factor evaluado	**Modelo A**	**Modelo B**
Tiempo de desarrollo	Medio-Bajo – 2-3 días	Medio – 5-7 días
Capacidad de actualización futura	Alta – Actualizaciones locales controladas	Alta – MCP adaptable a nuevas versiones de Claude o Wazuh
Riesgo de retrasos	Bajo – Arquitectura autónoma y local	Medio – Las actualizaciones o desconexiones pueden afectar a la continuidad del servicio

Tras realizar el análisis se puede concluir que ambos modelos presentan una viabilidad en la mayoría de los aspectos positiva, aunque con enfoques y fortalezas diferentes. El modelo A destaca por su simplicidad, bajo coste de implementación y completa soberanía de datos. No obstante, aunque es un entorno adecuado para defensa, al basarse únicamente en un script, es un modelo primitivo y sin interfaz. Esto también significa que es una solución con un margen de mejora considerable.

Por otro lado, el modelo B ofrece una mayor madurez tecnológica y modularidad, permitiendo una experimentación con IA más avanzada. Sin embargo, estas ventajas implican una mayor complejidad técnica y una limitación en la confidencialidad de los datos, al depender de servicios en la nube, lo que introduce consideraciones adicionales en materia de seguridad.

4.4.2. Marco legal y normativo

En el ámbito de la IA, el Reglamento (UE) 2024/1689 es el primer marco legal íntegramente europeo. Busca garantizar el buen uso y desarrollo de las herramientas de aprendizaje automático, y hace esto clasificando los sistemas en función del nivel de riesgo que suponen y estableciendo obligaciones proporcionales, especialmente en materia de ciberseguridad. Este reglamento impone a todos los sistemas de IA de alto riesgo, como es el caso de los sistemas de defensa, la obligación de incorporar trazabilidad, registro de operaciones y supervisión humana. Dentro del contexto de un SOC, esto se traduce en auditorías de las decisiones automáticas de la IA y en control de los algoritmos que priorizan las alertas y correlacionan eventos y que por tanto toman decisiones en el sistema.

También es importante incorporar registros de la documentación de las bases de datos empleadas, en su caso, para el proceso de entrenamiento, validación y pruebas del modelo utilizado, siguiendo los principios de transparencia y responsabilidad. [42]

Siguiendo la misma línea que el reglamento anteriormente citado, la Directiva (UE) 2022/2555, también exige que la aplicación de la IA esté acompañada de mecanismos de control y verificación humana, garantizando no comprometer la seguridad ni la responsabilidad del operador o analista. [43]

El Reglamento (UE) 2016/679, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales, permite excepcionalmente el uso de procesos automatizados de gran escala para el tratamiento de estos datos cuando existen motivos de seguridad nacional, teniendo la obligación de justificarlos bajo los principios de proporcionalidad y necesidad. En el caso de los proyectos que impliquen modelos de lenguaje avanzado en la detección de amenazas, estos deben diseñarse teniendo en cuenta además de su eficacia técnica, también el respeto a los derechos fundamentales y la protección de la información sensible.[44]

En el sector español, el Esquema Nacional de Seguridad (ENS), representa la principal referencia para las entidades que prestan servicio tecnológico al Estado. Dentro de este marco, se exigen unos requisitos mínimos para garantizar la confidencialidad, integridad y trazabilidad de los sistemas de información. El ENS proporciona una guía para la gestión de incidentes y la evaluación de riesgos, aplicable al contexto de un SOC, asegurando de esta manera la coherencia normativa con las directrices tanto españolas como europeas. También refuerza el principio de seguridad por diseño y por defecto, exigiendo que los sistemas tecnológicos incorporen medidas de protección desde su concepción, y las sigan aplicando durante todo su ciclo de vida. El Real Decreto 311/2021, que es el encargado de regular el ENS, también se encuentra alineado con las guías CCN-STIC, lo que facilita su aplicación en entornos de ciberdefensa. [45]

En conjunto, estas normas forman un entorno regulatorio robusto y coherente que favorece la seguridad y la privacidad, garantizando que las innovaciones en ciberdefensa se realicen bajo principios de seguridad jurídica y protección de datos En los entornos de los SOC, estas normativas implican que los modelos de IA integrados deben ser auditables, verificables y sometidos a supervisión humana. De esta forma se asegura además de la calidad del análisis, la transparencia en la toma de decisiones automatizadas.

4.4.3. Consideraciones de Seguridad

El uso de IA dentro de los ámbitos de defensa requiere un análisis riguroso de la seguridad, tanto de la naturaleza de los datos procesados como de los entornos donde se desplieguen los modelos.

Es importante conocer la naturaleza de la información procesada, no todos los datos tienen el mismo valor ni el mismo riesgo, por lo que antes de definir una infraestructura, hay que entender qué clase de datos va a tratar la IA o el SIEM. Esto implica analizar exhaustivamente diferentes aspectos:

En primer lugar, la clasificación y sensibilidad de la información, en este caso los logs y alertas pueden revelar patrones de actividad interna, la topología de la red o incluso vulnerabilidades de ésta. En segundo lugar, el origen y trazabilidad de los datos, los cuales proceden de sensores o agentes de Wazuh, y, por último, la finalidad del tratamiento, donde entra en juego la seguridad que la IA tiene, ya que es esta quien trata estos datos o alertas, por lo que requiere un nivel de protección adicional. Si los datos incluyen identificadores personales o direcciones IPs militares, su exposición en la nube vulnera políticas de seguridad y normativas.

Cuando hablamos de la arquitectura de despliegue del modelo de IA, es de obligación decidir entre infraestructura local o infraestructura en la nube. Esta elección implica diferencias sustanciales en los distintos niveles de control y exposición en la red.

Teniendo en cuenta la naturaleza de la información procesada, descrita anteriormente, los despliegues en la nube, aunque aporten elasticidad y escalabilidad, gracias a su capacidad de cómputo flexible también introducen un riesgo potencial de exfiltración de la información o tratamiento no autorizado de información sensible.

En cambio, las infraestructuras locales permiten tener una custodia y control total sobre los datos, ya que estos se encuentran en la propia arquitectura de la organización. Esto reduce significativamente la superficie de ataque y minimiza la exposición frente a amenazas externas.

Se ha considerado conveniente estudiar la posibilidad de securizar la IA. La manera más adecuada es securizar la IA del modelo experimental, debido a que los contenedores (Docker) permiten un aislamiento lógico de la IA dentro del sistema anfitrión. De acuerdo con las recomendaciones del CCN, este aislamiento se debe completar con un conjunto de medidas específicas y con un modelo de Ollama. [46] En primer lugar, se debe ejecutar el contenedor de la IA en una red interna cerrada, restringiendo la comunicación únicamente con MCP Server. En segundo lugar, se recomienda incorporar un sistema de filtrado y validación de prompts, mediante el módulo Pipelines de OpenWeb UI. Con este componente, se inspeccionan las entradas que llegan al modelo, evitando inyecciones de prompts o intentos de manipulación de contexto. En tercer lugar, de manera complementaria se puede instalar un modelo de moderación llamado Llama Guard, que detecta las entradas y salidas, para detectar información sensible o comando potencialmente dañinos. Por último, otro aspecto clave es mantener un registro actualizado de los componentes instalados mediante un SBOM[18], permitiendo auditar las versiones y dependencias utilizadas, mejorando su ciberseguridad detectando de manera temprana vulnerabilidades o incumplimientos de estándares de seguridad.

5. CONCLUSIONES

Este Trabajo de Fin de Grado ha buscado analizar y demostrar la viabilidad para integrar sistemas de IA en entornos de ciberdefensa militar, concretamente en los SOC que hacen uso de plataformas SIEM. Este estudio se ha llevado a cabo mediante la aplicación de una metodología de investigación aplicada, llevando a cabo una revisión bibliográfica, el diseño de modelos de integración, su implementación en un entorno controlado y la evaluación de sus viabilidades. Se ha trabajado con herramientas de código abierto y de despliegue local tanto para el entorno como para la plataforma SIEM y los modelos de IA (Wazuh, Ollama, Docker), y con un modelo en la nube de software comercial (Claude), permitiendo así un análisis más completo de las diferentes posibilidades de integración, considerando aspectos de seguridad y rendimiento.

A partir de la revisión bibliográfica, se ha podido comprobar que la ciberdefensa constituye un nuevo dominio estratégico, caracterizado por su anonimato, que lo constituye un entorno asimétrico en el que tanto actores estatales como no estatales pueden ser factores de riesgo para la estabilidad de un Estado. Los principales actores y amenazas evidencian la diversidad existente en el panorama actual, con casos que han puesto de manifiesto la necesidad de reforzar la resiliencia tecnológica. Dentro del ámbito estrictamente militar se ha podido constatar que existe una tendencia a estructuras modulares y autónomas que son capaces de operar en cualquier teatro de operaciones.

[18] El Software Bill of Material es un documento que actúa como una lista de los componentes, bibliotecas y dependencias que integran un software. Su propósito es mejorar la transparencia y trazabilidad de la cadena de suministro. Fuente: https://www.cisa.gov/sbom

La aplicación de la IA en este marco no es solo una tendencia tecnológica, sino una necesidad estratégica, que refuerza la capacidad de los SOC y se está consolidando como un elemento diferenciador en la gestión de alertas y activos. Su despliegue plantea desafíos técnicos y éticos, destacando la dependencia tecnológica, la trazabilidad y sensibilidad de la información y la adecuación a la normativa vigente.

El análisis de necesidades y el estudio DAFO han permitido identificar y corroborar con especialistas estos aspectos, identificando las diferentes fortalezas, debilidades, amenazas y oportunidades de la integración SIEM-IA.

Mediante una comparativa de prestaciones y, de manera complementaria, la metodología AHP, se han evaluado y seleccionado los distintos modelos de IA y plataformas SIEM, teniendo en cuenta factores de despliegue, coste, integración, rendimiento y soporte. En el ámbito de la IA, se seleccionó Mistral como modelo más adecuado, gracias a su equilibrio entre herramienta de código abierto y rendimiento, así como su despliegue local. Como modelo de contraste, se vio adecuado seleccionar una herramienta que trabajase en la nube y dentro de la arquitectura MCP, como es Claude. Respecto a las plataformas SIEM, se ha seleccionado Wazuh, siendo esta la más idónea por su compatibilidad con modelos de IA y equilibrio entre rendimiento y soberanía de datos, otras soluciones, pese a su madurez y funcionalidades avanzadas, han resultado menos adecuadas por su elevado coste y dependencia de servicios en la nube.

En la parte experimental del trabajo, se desarrollaron tres modelos de integración. El modelo práctico, basado en la API de Wazuh Indexer y el modelo Mistral, ha demostrado ser el más funcional, consiguiendo una comunicación directa y estable para el análisis automatizado de las alertas. Por otro lado, el modelo experimental con MCP Server, evidenció potencial para arquitecturas más complejas, pero también ha tenido limitaciones derivadas de incompatibilidades técnicas. Estos resultados validan la posibilidad real de integrar IA en entornos SOC militares, pudiendo servir de base de estudio para futuros proyectos.

El análisis de viabilidad mediante la metodología TELOS, permitió medir las dimensiones más importantes de los modelos práctico y experimental. El modelo práctico alcanzó una viabilidad alta, gracias al empleo de software libre y su autonomía. Por otro lado, el modelo experimental resultó más modular y con mayor capacidad para implementar securización, pero con mayor complejidad técnica y menos soberanía de datos. El estudio concluye que los modelos más coherentes con las necesidades de ciberdefensa son aquellos cuyas prestaciones son locales, tanto por seguridad como por sostenibilidad operativa.

En conjunto, se han alcanzado los objetivos planteados: el análisis documental ha permitido identificar los principales retos tecnológicos de la IA en ciberdefensa; la selección y evaluación de herramientas evidenció la superioridad de las soluciones locales y de código abierto; y el diseño validó la viabilidad práctica de la integración propuesta.

Las líneas futuras de trabajo, provenientes del análisis de los resultados y limitaciones, son tres: la estandarización de las comunicaciones entre SIEM e IA, con el impulso del uso de protocolos MCP y APIs seguras para lograr la interoperabilidad total; la securización más avanzada de los modelos locales; y el diseño de una interfaz gráfica que sustituya el script experimental del modelo práctico y facilite el uso operativo en entornos reales.

Finalmente, se propone continuar con esta línea de investigación hacia la creación de SOC inteligentes con capacidades predictivas y de aprendizaje automático y profundo, fortaleciendo de esta manera la ciberdefensa nacional.

REFERENCIAS BIBLIOGRÁFICAS

[1] Escuela de Altos Estudios de la Defensa de España, "Estrategia de la información y seguridad en el ciberespacio," *Estrateg. la Inf. y Segur. en el ciberespacio*, p. 125, 2014, Accessed: Sep. 11, 2025. [Online]. Available: https://dialnet.unirioja.es/servlet/libro?codigo=559597&info=resumen&idioma=SPA

[2] "Estrategia de Seguridad Nacional - Ministerio de Defensa de España." Accessed: Sep. 30, 2025. [Online]. Available: https://www.defensa.gob.es/defensa/politicadefensa/estrategiaseguridad/

[3] L. Rouhiainen, "Inteligencia artificial," *Madrid Alienta Editor.*, pp. 20–21, 2018.

[4] J. L. Escrivá, "Estrategia de Inteligencia Artificial 2024," in *Gobierno de España*, 2024, pp. 6–20. Accessed: Sep. 11, 2025. [Online]. Available: https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf

[5] A. Dolores and Z. Rendón, "Impacto de la inteligencia artificial en los ciberataques," *Sinapsis La Rev. científica del ITSUP, ISSN-e 1390-9770, Vol. 24, Nº. 1, 2024 (Ejemplar Dedic. a La Educ. Tecnológica Contrib. en la Investig. Científica)*, vol. 24, no. 1, p. 6, 2024, Accessed: Sep. 11, 2025. [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=9642443&info=resumen&idioma=SPA

[6] A. Felipe and H. Salazar, "Experiencia en el uso de la ia aplicada a un centro de operaciones de ciberseguridad para la empresa Emergia," 2025, *Universidad Católica de Manizales*. Accessed: Sep. 11, 2025. [Online]. Available: https://repositorio.ucm.edu.co/handle/10839/4641

[7] F. A. Marín Gutiérrez, "La inteligencia artificial en el campo de la informació: su utilización en apoyo a la desinformación," *Usos Mil. la Intel. Artif. la Autom. y la robótica (IAA&R), 2020, págs. 69-96*, pp. 69–96, 2020, Accessed: Sep. 16, 2025. [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=7771638&info=resumen&idioma=ENG

[8] S. Tulasi Prasad, "Ethical Hacking and Types of Hackers," *Int. J. Emerg. Technol. Comput. Sci. Electron.*.

[9] C. Alberto Flores Quispe, "TIPOS DE HACKERS", Accessed: Sep. 18, 2025. [Online]. Available: http://www.axelsanmiguel.com

[10] Instituto Nacional de Ciberseguridad, "Guía de ciberataques: Todo lo que debes saber a nivel usuario," Oct. 2020. Accessed: Sep. 18, 2025. [Online]. Available: https://www.incibe.es/ciudadania/formacion/guias/guia-de-ciberataques

[11] "¿Qué es el ataque de phishing? Tipos y ejemplos — Wallarm." Accessed: Sep. 23, 2025. [Online]. Available: https://lab.wallarm.com/what/ataque-de-phishing-🎯/?lang=es

[12] O. I. Falowo, S. Popoola, J. Riep, V. A. Adewopo, and J. Koch, "Threat Actors' Tenacity to Disrupt: Examination of Major Cybersecurity Incidents," *IEEE Access*, vol. 10, pp. 134038–134051, 2022, doi: 10.1109/ACCESS.2022.3231847.

[13] "Todo sobre ransomware: guía básica y preguntas frecuentes." Accessed: Sep. 23, 2025. [Online]. Available: https://www.welivesecurity.com/la-es/2014/06/10/todo-sobre-ransomware-guia-basica-preguntas-frecuentes/

[14] R. Ottis, "Analysis of the 2007 Cyber Attacks Against Estonia from the Information Warfare Perspective," in *Proceedings of the 7th European Conference on Information Warfare and Security*, Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE), 2008, pp. 163–168. [Online]. Available: https://ccdcoe.org/library/publications/analysis-of-the-2007-cyber-attacks-against-estonia-from-the-information-warfare-perspective/

[15] A. Kazi, S. Kazi, and S. Bhosale, “Invisible Battlefields: Analyzing the Viasat Attack and its Broader Implications,” *Sci. Bull.*, vol. 30, no. 1, pp. 59–67, Jun. 2025, doi: 10.2478/BSAFT-2025-0007.

[16] A. Perdana, M. E. Aminanto, and B. Anggorojati, “Hack, heist, and havoc: The Lazarus Group’s triple threat to global cybersecurity,” *J. Inf. Technol. Teach. Cases*, 2024, doi: 10.1177/20438869241303941.

[17] D. E. Cabuya-Padilla, & Carlos, and A. Castaneda-Marroquin, “Marco de referencia para el modelamiento y simulación de la ciberdefensa marítima - MARCIM: estado del arte y metodología,” *DYNA Rev. la Fac. Minas. Univ. Nac. Colomb. Sede Medellín, ISSN 0012-7353, Vol. 91, Nº. 231, 2024, págs. 169-179*, vol. 91, no. 231, pp. 169–179, 2024, doi: 10.15446/dyna.v91n231.109774.

[18] C. Baselga Mateo and L. Martínez Saura, “La ciberdefensa táctica,” *Ejército Rev. del Ejército Tierra español*, no. 962, pp. 88–93, Jun. 2021, [Online]. Available: https://publicaciones.defensa.gob.es/

[19] G. González-Granadillo, S. González-Zarzosa, and R. Diaz, “Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures,” *Sensors 2021, Vol. 21, Page 4759*, vol. 21, no. 14, p. 4759, Jul. 2021, doi: 10.3390/S21144759.

[20] R. López, D. E. Mántaras, and Y. Pere Brunet, “¿Qué es la inteligencia artificial?,” *Papeles Relac. ecosociales y cambio Glob. ISSN 1888-0576, Nº. 164 (Riesgos, Vent. y Reper. la Intel. Artif. 2023 (Ejemplar Dedic. a Intel. Artif. págs. 13-21*, no. 164, pp. 13–21, 2023, Accessed: Sep. 11, 2025. [Online]. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=9287111&info=resumen&idioma=ENG

[21] I. Gartner, “Magic Quadrant for Security Information and Event Management,” 2024. [Online]. Available: https://www.gartner.com/doc/reprints?id=1-2F6JB0XP&ct=231002&st=sb

[22] “Aplicación QRadar Advisor with Watson - Documentación de IBM.” Accessed: Sep. 25, 2025. [Online]. Available: https://www.ibm.com/docs/es/qradar-common?topic=apps-qradar-advisor-watson-app

[23] “QRadar Advisor with Watson 2.6 brings AI-powered investigations to Analyst Workflow UI.” Accessed: Sep. 25, 2025. [Online]. Available: https://community.ibm.com/community/user/blogs/jo-leger1/2020/10/16/qradar-advisor-with-watson-26-brings-ai-powered-in

[24] “Security Onion: Security Onion 2.4.110 now available including new AI Summary feature and much more!” Accessed: Sep. 25, 2025. [Online]. Available: https://blog.securityonion.net/2024/10/security-onion-24110-hurricane-helene.html

[25] F. Linares-Torres, “Inteligencia Artificial y Ciberdefensa,” *Rev. Segur. y Pod. Terr.*, vol. 3, no. 4, pp. 139–150, 2024, doi: 10.56221/SPT.V3I4.72.

[26] F. Linares-Torres, “Inteligencia artificial, desinformación y defensa nacional,” *Rev. la Esc. Super. Guerr. Nav.*, vol. 21, no. 2, pp. 58–71, Dec. 2024, doi: 10.35628/resup.v16i2.149.

[27] “The EU’s Cybersecurity Strategy for the Digital Decade | Shaping Europe’s digital future.” Accessed: Sep. 29, 2025. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/eus-cybersecurity-strategy-digital-decade-0

[28] “España Digital | España Digital 2026.” Accessed: Sep. 29, 2025. [Online]. Available: https://espanadigital.gob.es/espana-digital

[29] “mistral.” Accessed: Oct. 02, 2025. [Online]. Available: https://ollama.com/library/mistral

[30] “Meta Llama 3 License.” Accessed: Sep. 30, 2025. [Online]. Available:

https://www.llama.com/llama3/license/?utm_source=chatgpt.com

[31] "Investigación de OpenAI | OpenAI." Accessed: Oct. 09, 2025. [Online]. Available: https://openai.com/es-ES/research/index/

[32] "Overview | Claude." Accessed: Oct. 09, 2025. [Online]. Available: https://claude.com/product/overview

[33] "DeepSeek." Accessed: Oct. 09, 2025. [Online]. Available: https://www.deepseek.com/en

[34] "User manual · Wazuh documentation." Accessed: Oct. 02, 2025. [Online]. Available: https://documentation.wazuh.com/current/user-manual/

[35] "Security Onion Solutions." Accessed: Oct. 09, 2025. [Online]. Available: https://securityonionsolutions.com/software

[36] "The Splunk Platform | Splunk." Accessed: Oct. 09, 2025. [Online]. Available: https://www.splunk.com/en_us/products/platform.html

[37] "IBM QRadar SIEM." Accessed: Oct. 09, 2025. [Online]. Available: https://www.ibm.com/es-es/products/qradar-siem

[38] "¿Qué es SIEM de Microsoft Sentinel? | Microsoft Learn." Accessed: Oct. 09, 2025. [Online]. Available: https://learn.microsoft.com/es-es/azure/sentinel/overview?tabs=defender-portal

[39] "CCN-CERT ." Accessed: Oct. 02, 2025. [Online]. Available: https://www.ccn-cert.cni.es/es/

[40] "GitHub - gbrigandi/mcp-server-wazuh: MCP Server for Wazuh SIEM." Accessed: Oct. 06, 2025. [Online]. Available: https://github.com/gbrigandi/mcp-server-wazuh

[41] European Knowledge Center for Information Technology, "Estudio de viabilidad de un proyecto de software," Sep. 2022, *TIC Portal*. [Online]. Available: https://www.ticportal.es/glosario-tic/estudio-viabilidad-software

[42] O. de Publicaciones de la Unión Europea and L. Luxemburgo, "Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.° 300/2008, (UE) n.° 167/2013, (UE) n.° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial)Texto pertinente a efectos del EEE." [Online]. Available: http://data.europa.eu/eli/reg/2024/1689/oj

[43] Parlamento Europeo y Consejo de la Unión Europea, "Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión (Directiva SRI 2)," Dec. 2022, *Publications Office of the European Union*. [Online]. Available: https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32022L2555

[44] Parlamento Europeo y Consejo de la Unión Europea, "Reglamento (Ue) 2016/679 Del Parlamento Europeo Y Del Consejo De 27 De Abril De 2016," *D. Of. la Unión Eur.*, 2016.

[45] Ministerio de la Presidencia Relaciones con las Cortes y Memoria Democrática, "Real Decreto 43/2021, de 26 de enero, por el que se desarrolla el Real Decreto-ley 12/2018, de 7 de septiembre, de seguridad de las redes y sistemas de información," Oct. 2021. [Online]. Available: https://www.boe.es/buscar/doc.php?id=BOE-A-2021-1192

[46] Centro Criptológico Nacional (CCN), "CCN-TEC 014: Instalación y Securización de un ChatBot con LLM de forma local," Madrid, España, Mar. 2025. [Online]. Available: https://www.ccn.cni.es

Fuentes orales

Cuestionarios de valoración:

Título: “Cuestionario de valoración sobre integración de IA en plataformas SIEM”.

Fecha de aplicación: 24 – 26 de octubre de 2025.

Participantes: 10 especialistas en ciberseguridad (7 analistas, 1 especialista en I+D, 2 jefes de SOC)

Método: cuestionario semiestructurado con preguntas tipo Likert (1-5).

Modalidad: telemática.

Objetivo: complementar el análisis DAFO y contrastar percepciones sobre debilidades y fortalezas de la integración SIEM-IA.

Referencia: Ver Anexo I para el cuestionario completo.

Anexo I. Diagrama de Gantt

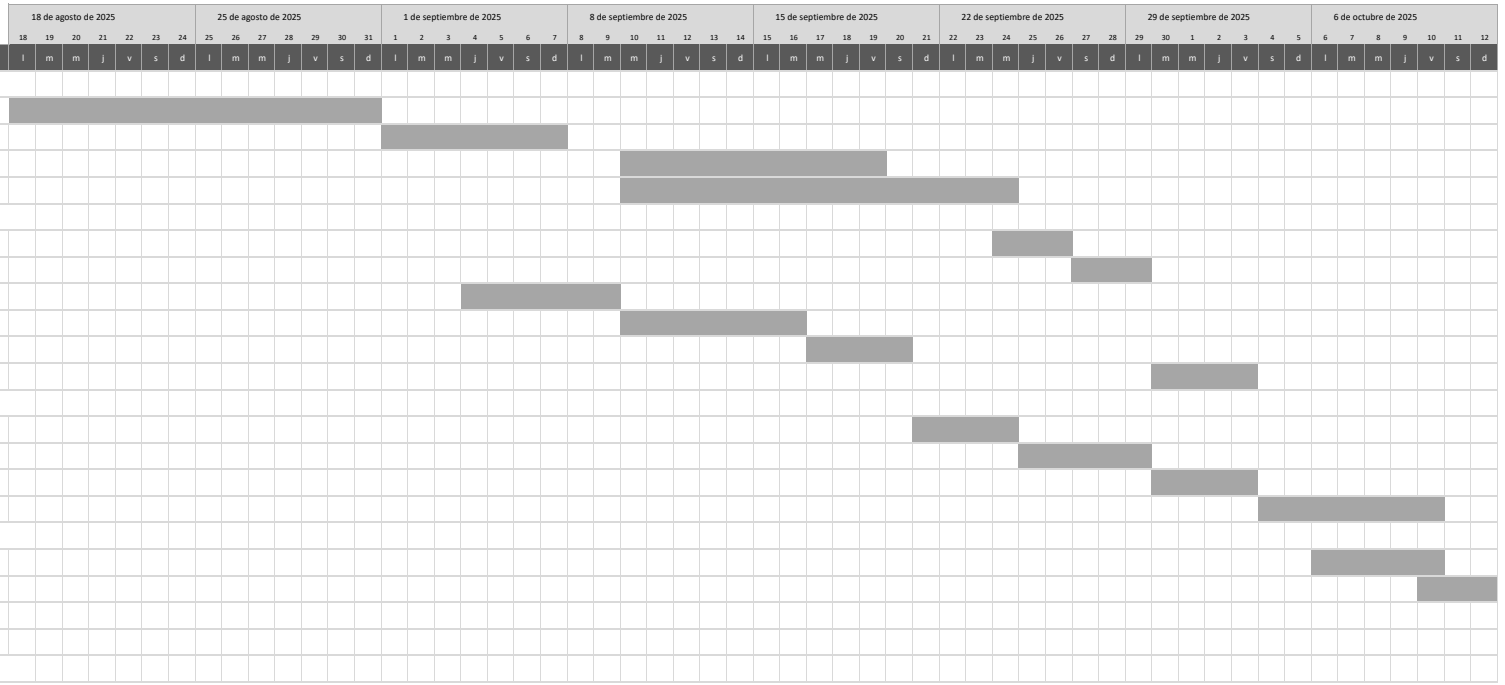
TFG

Juan Carlos Morales Rico

Inicio del proyecto: lu, 2025-08-18

Semana para mostrar: 1

TAREA	PROGRESO	INICIO	FIN
Investigación Inicial			
Revisión de conceptos teóricos	100%	18-8-25	31-8-25
Estudio del estado del arte: casos de uso y sistemas actuales	100%	1-9-25	7-9-25
Recopilación bibliográfica y normativa para marco teórico	100%	10-9-25	19-9-25
Elaboración del marco teórico	100%	10-9-25	24-9-25
Análisis y Planificación			
Redacción Análisis de necesidades en ciberdefensa	100%	24-9-25	26-9-25
Redacción Justificación técnica	100%	27-9-25	29-9-25
Selección de herramientas y tecnologías	100%	4-9-25	9-9-25
Testeo de soluciones y entorno	100%	10-9-25	16-9-25
Planificación de la arquitectura	100%	17-9-25	20-9-25
Redacción de la fase en el Desarrollo del TFG	100%	30-9-25	3-10-25
Diseño Preliminar			
Diseño de la arquitectura del sistema propuesto	100%	21-9-25	24-9-25
Integración con infraestructuras existentes	100%	25-9-25	29-9-25
Simulación o prototipado	100%	30-9-25	3-10-25
Redacción de la Fase en el Desarrollo del TFG	100%	4-10-25	10-10-25
Evaluación y Viabilidad			
Pruebas del prototipo y análisis de resultados	100%	6-10-25	10-10-25
Evaluación de viabilidad técnica y económica	100%	10-10-25	13-10-25
Marco legal y normativo	100%	15-10-25	16-10-25
Evaluación de seguridad	100%	17-10-25	17-10-25



Fuente: https://create.microsoft.com/es-es/template/diagrama-de-gantt-simple-4bf6b793-490f-4623-84ca-c9c6251a91fc

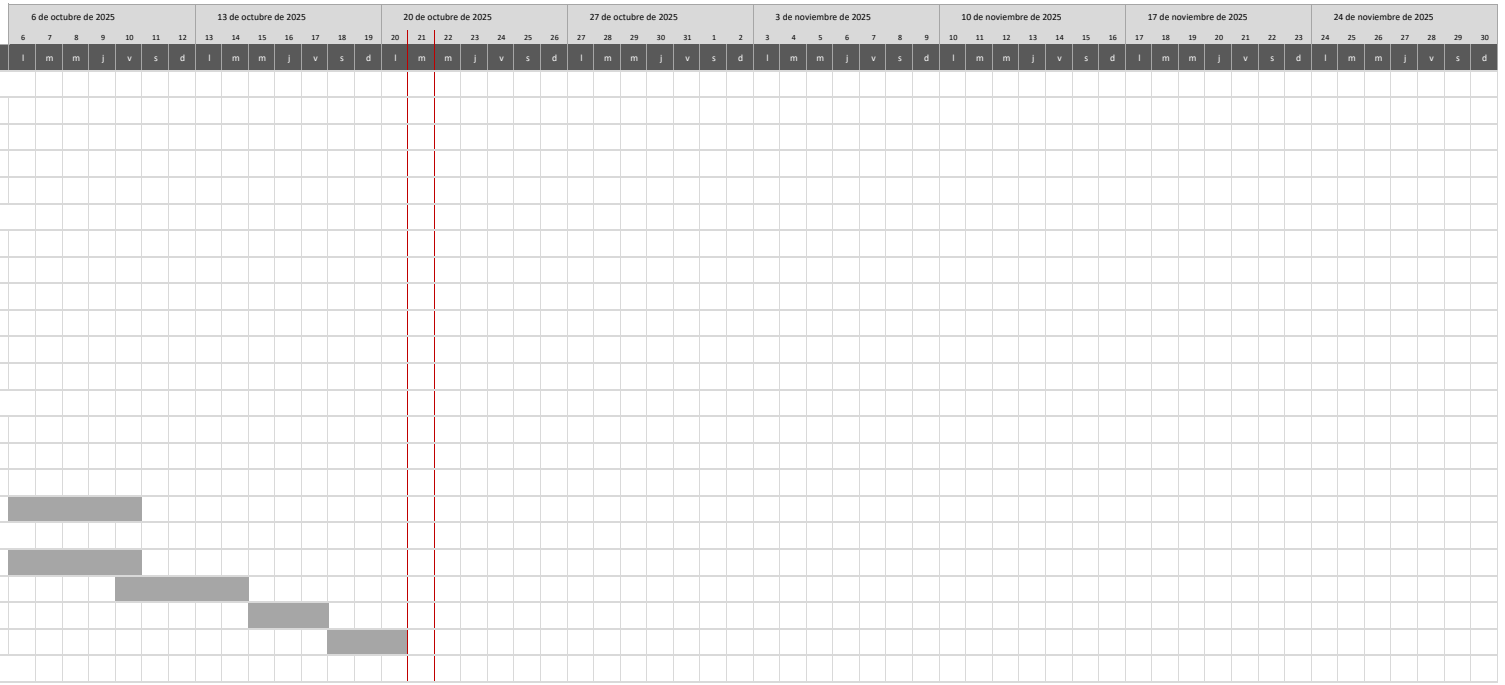
TFG

Juan Carlos Morales Rico

Inicio del proyecto: lu, 2025-08-18

Semana para mostrar: 8

TAREA	PROGRESO	INICIO	FIN
Investigación Inicial			
Revisión de conceptos teóricos	100%	18-8-25	31-8-25
Estudio del estado del arte: casos de uso y sistemas actuales	100%	1-9-25	7-9-25
Recopilación bibliográfica y normativa para marco teórico	100%	10-9-25	19-9-25
Elaboración del marco teórico	100%	10-9-25	24-9-25
Análisis y Planificación			
Redacción Análisis de necesidades en ciberdefensa	100%	24-9-25	26-9-25
Redacción Justificación técnica	100%	27-9-25	29-9-25
Selección de herramientas y tecnologías	100%	4-9-25	9-9-25
Testeo de soluciones y entorno	100%	10-9-25	16-9-25
Planificación de la arquitectura	100%	17-9-25	20-9-25
Redacción de la fase en el Desarrollo del TFG	100%	30-9-25	3-10-25
Diseño Preliminar			
Diseño de la arquitectura del sistema propuesto	100%	21-9-25	24-9-25
Integración con infraestructuras existentes	100%	25-9-25	29-9-25
Simulación o prototipado	100%	30-9-25	3-10-25
Redacción de la Fase en el Desarrollo del TFG	100%	4-10-25	10-10-25
Evaluación y Viabilidad			
Pruebas del prototipo y análisis de resultados	100%	6-10-25	10-10-25
Evaluación de viabilidad técnica y económica	100%	10-10-25	14-10-25
Marco legal y normativo	100%	15-10-25	17-10-25
Evaluación de seguridad	100%	18-10-25	20-10-25



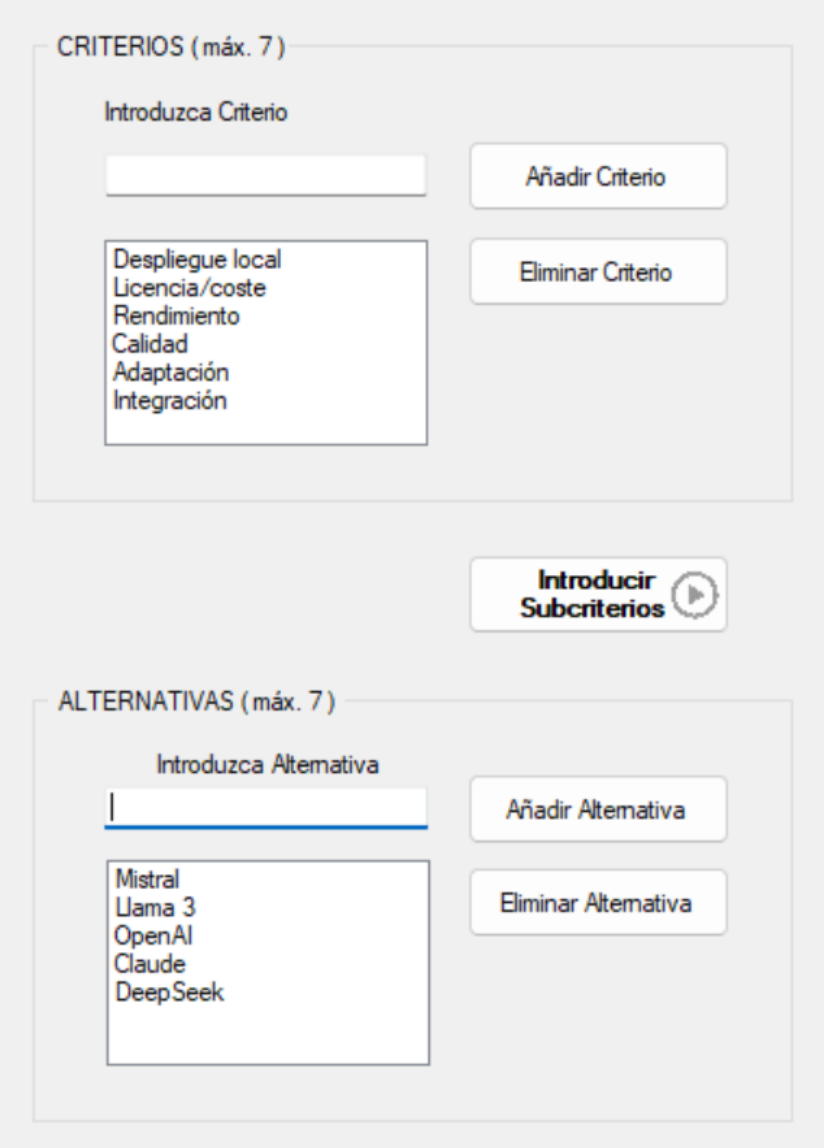
Anexo II. Preguntas cuestionario de valoración

1. *La integración de IA en los SIEM mejora la detección de amenazas*
2. *El uso de soluciones basadas en la nube reduce la soberanía de datos*
3. *La IA puede contribuir a la gestión de alertas y reducción de falsos positivos*
4. *Considera que existe un marco normativo suficiente para el uso de IA en Ciberdefensa*
5. *En ciberdefensa es mejor el software de código abierto que los softwares comerciales*
6. *Los SIEM utilizados actualmente corren riesgo de quedarse obsoletos*
7. *El uso de IA por parte de los ciberatacantes aumenta la carga de trabajo de un SOC*

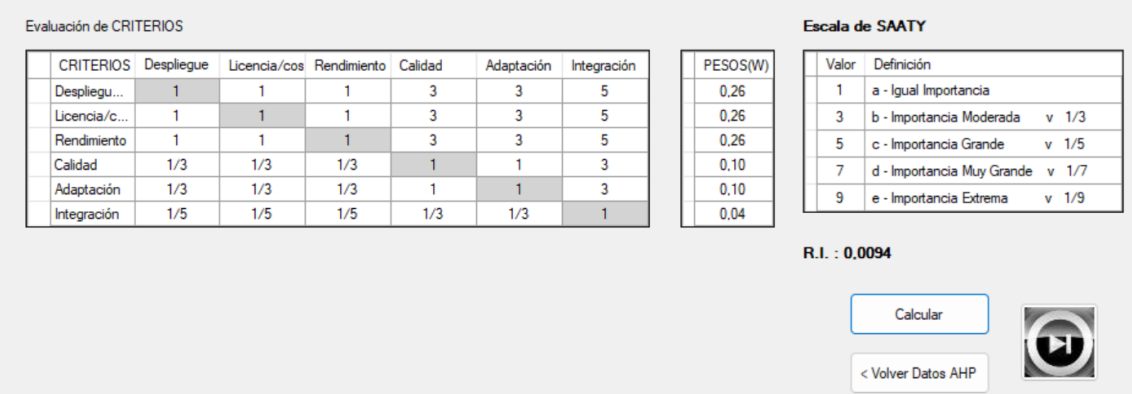
Anexo III. Desarrollo del análisis AHP para soluciones de IA

En este anexo, se muestran los resultados obtenidos durante el desarrollo del análisis AHP aplicado a las diferentes soluciones IA que se contemplan en el trabajo. Las figuras siguientes muestran los criterios, comparaciones y resultados de prioridades calculados en el proceso.

En la siguiente figura se muestran los criterios principales y las alternativas que han sido definidas.



Se muestra la matriz de comparación por pares entre los criterios previamente definidos, de esta forma se obtiene la ponderación relativa de cada uno.



Evaluación de CRITERIOS

CRITERIOS	Despliegue	Licencia/cos	Rendimiento	Calidad	Adaptación	Integración
Despliegu...	1	1	1	3	3	5
Licencia/c...	1	1	1	3	3	5
Rendimiento	1	1	1	3	3	5
Calidad	1/3	1/3	1/3	1	1	3
Adaptación	1/3	1/3	1/3	1	1	3
Integración	1/5	1/5	1/5	1/3	1/3	1

PESOS(W)
0,26
0,26
0,26
0,10
0,10
0,04

Escala de SAATY

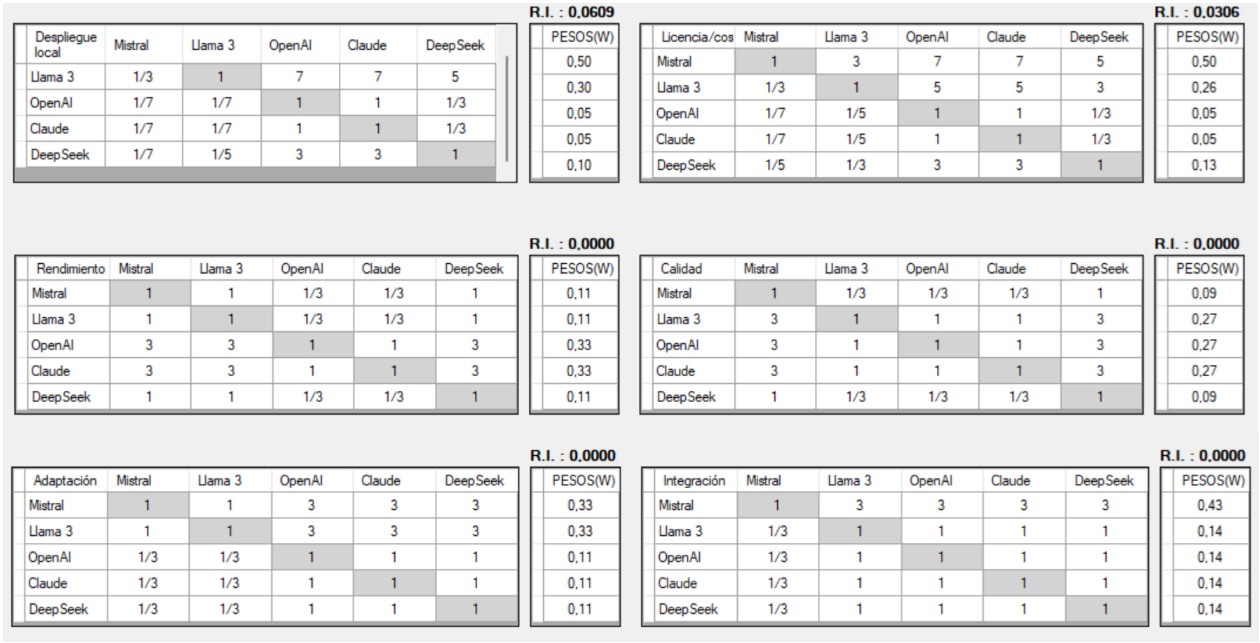
Valor	Definición
1	a - Igual Importancia
3	b - Importancia Moderada v 1/3
5	c - Importancia Grande v 1/5
7	d - Importancia Muy Grande v 1/7
9	e - Importancia Extrema v 1/9

R.I. : 0,0094

Calcular

< Volver Datos AHP

Por último, a continuación, se muestra el análisis de las alternativas en función de los criterios establecidos. Una tabla por cada criterio, en cada celda de cada tabla se refleja la valoración relativa de los modelos de IA por pares.



R.I. : 0,0609

Despliegue local	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Llama 3	1/3	1	7	7	5
OpenAI	1/7	1/7	1	1	1/3
Claude	1/7	1/7	1	1	1/3
DeepSeek	1/7	1/5	3	3	1

PESOS(W)
0,50
0,30
0,05
0,05
0,10

R.I. : 0,0306

Licencia/cos	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Mistral	1	3	7	7	5
Llama 3	1/3	1	5	5	3
OpenAI	1/7	1/5	1	1	1/3
Claude	1/7	1/5	1	1	1/3
DeepSeek	1/5	1/3	3	3	1

PESOS(W)
0,50
0,26
0,05
0,05
0,13

R.I. : 0,0000

Rendimiento	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Mistral	1	1	1/3	1/3	1
Llama 3	1	1	1/3	1/3	1
OpenAI	3	3	1	1	3
Claude	3	3	1	1	3
DeepSeek	1	1	1/3	1/3	1

PESOS(W)
0,11
0,11
0,33
0,33
0,11

R.I. : 0,0000

Calidad	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Mistral	1	1/3	1/3	1/3	1
Llama 3	3	1	1	1	3
OpenAI	3	1	1	1	3
Claude	3	1	1	1	3
DeepSeek	1	1/3	1/3	1/3	1

PESOS(W)
0,09
0,27
0,27
0,27
0,09

R.I. : 0,0000

Adaptación	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Mistral	1	1	3	3	3
Llama 3	1	1	3	3	3
OpenAI	1/3	1/3	1	1	1
Claude	1/3	1/3	1	1	1
DeepSeek	1/3	1/3	1	1	1

PESOS(W)
0,33
0,33
0,11
0,11
0,11

R.I. : 0,0000

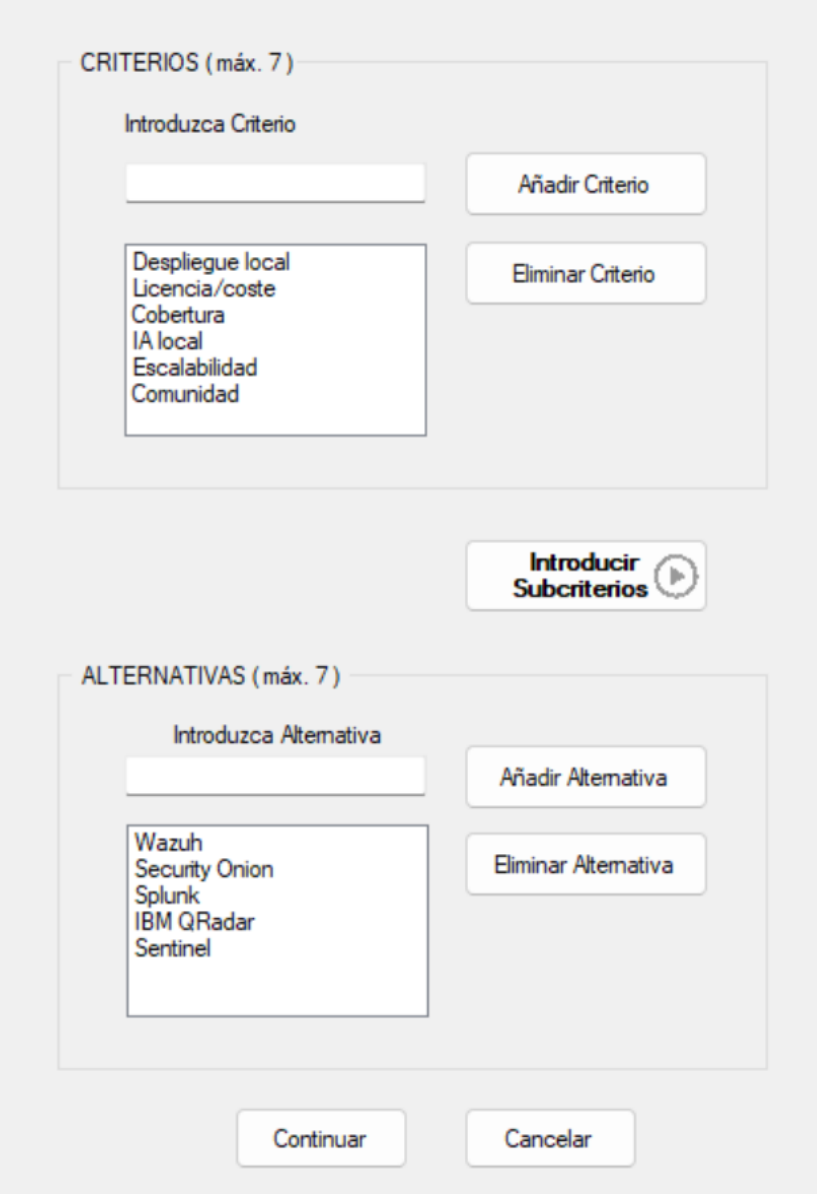
Integración	Mistral	Llama 3	OpenAI	Claude	DeepSeek
Mistral	1	3	3	3	3
Llama 3	1/3	1	1	1	1
OpenAI	1/3	1	1	1	1
Claude	1/3	1	1	1	1
DeepSeek	1/3	1	1	1	1

PESOS(W)
0,43
0,14
0,14
0,14
0,14

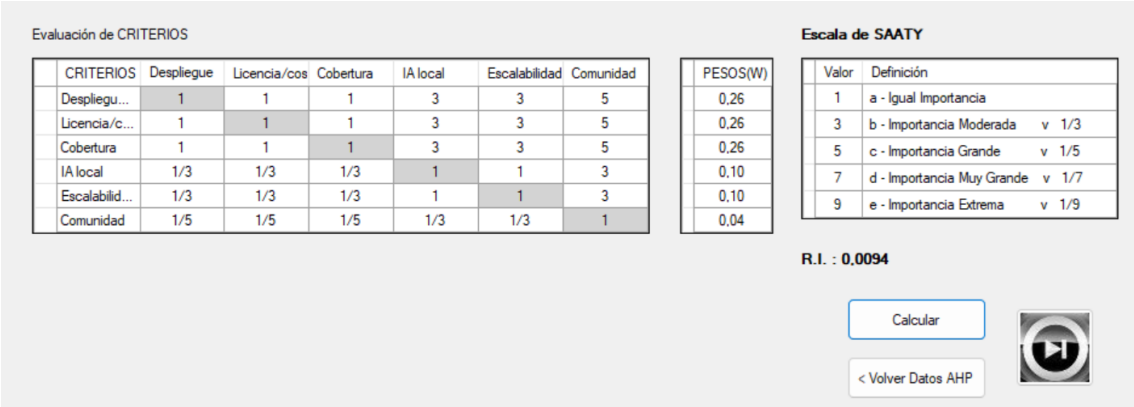
Anexo IV. Desarrollo del análisis AHP para soluciones SIEM

En este anexo, se muestran los resultados obtenidos durante el desarrollo del análisis AHP aplicado a las diferentes soluciones SIEM que se contemplan en el trabajo. Las figuras siguientes muestran los criterios, comparaciones y resultados de prioridades calculados en el proceso.

En la siguiente figura se muestran los criterios principales y las alternativas que han sido definidas.



Se muestra la matriz de comparación por pares entre los criterios previamente definidos, de esta forma se obtiene la ponderación relativa de cada uno.



Evaluación de CRITERIOS

CRITERIOS	Despliegue	Licencia/cos	Cobertura	IA local	Escalabilidad	Comunidad	PESOS(W)
Despliegu...	1	1	1	3	3	5	0,26
Licencia/c...	1	1	1	3	3	5	0,26
Cobertura	1	1	1	3	3	5	0,26
IA local	1/3	1/3	1/3	1	1	3	0,10
Escalabilid...	1/3	1/3	1/3	1	1	3	0,10
Comunidad	1/5	1/5	1/5	1/3	1/3	1	0,04

Escala de SAATY

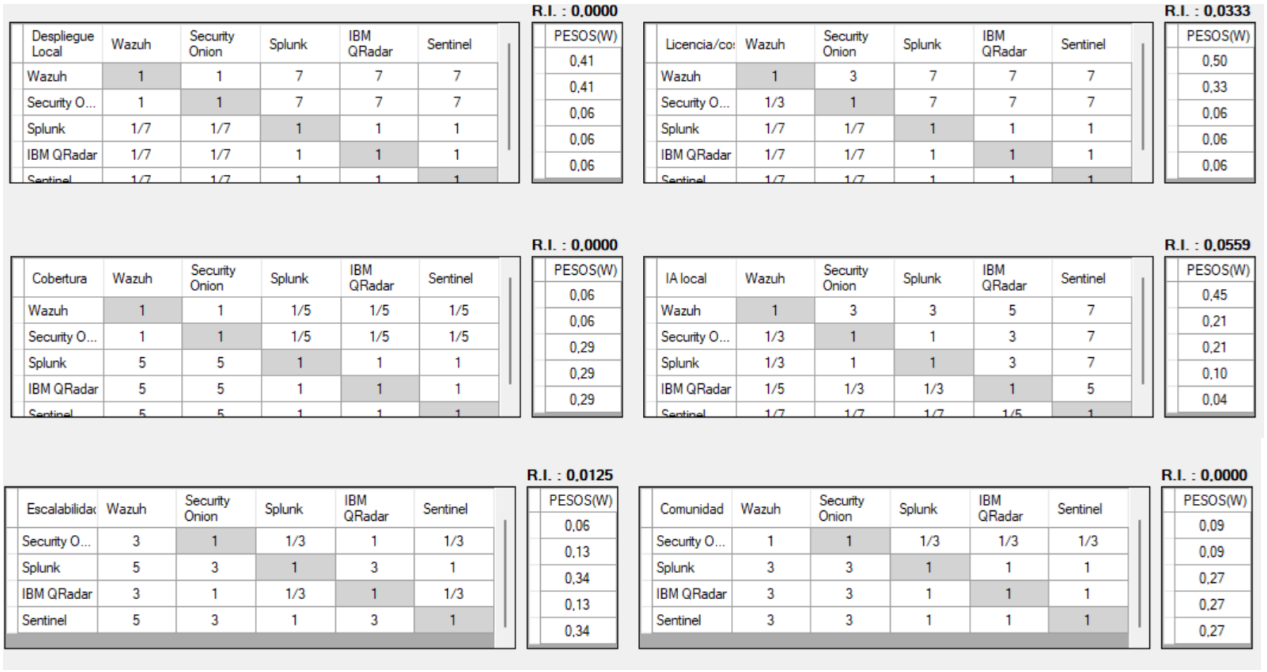
Valor	Definición
1	a - Igual Importancia
3	b - Importancia Moderada v 1/3
5	c - Importancia Grande v 1/5
7	d - Importancia Muy Grande v 1/7
9	e - Importancia Extrema v 1/9

R.I. : 0,0094

Calcular

< Volver Datos AHP

Por último, a continuación, se muestra el análisis de las alternativas en función de los criterios establecidos. Una tabla por cada criterio, en cada celda de cada tabla se refleja la valoración relativa de las plataformas SIEM por pares.



R.I. : 0,0000

Despliegue Local	Wazuh	Security Onion	Splunk	IBM QRadar	Sentinel	PESOS(W)
Wazuh	1	1	7	7	7	0,41
Security O...	1	1	7	7	7	0,41
Splunk	1/7	1/7	1	1	1	0,06
IBM QRadar	1/7	1/7	1	1	1	0,06
Sentinel	1/7	1/7	1	1	1	0,06

R.I. : 0,0333

Licencia/co	Wazuh	Security Onion	Splunk	IBM QRadar	Sentinel	PESOS(W)
Wazuh	1	3	7	7	7	0,50
Security O...	1/3	1	7	7	7	0,33
Splunk	1/7	1/7	1	1	1	0,06
IBM QRadar	1/7	1/7	1	1	1	0,06
Sentinel	1/7	1/7	1	1	1	0,06

R.I. : 0,0000

Cobertura	Wazuh	Security Onion	Splunk	IBM QRadar	Sentinel	PESOS(W)
Wazuh	1	1	1/5	1/5	1/5	0,06
Security O...	1	1	1/5	1/5	1/5	0,06
Splunk	5	5	1	1	1	0,29
IBM QRadar	5	5	1	1	1	0,29
Sentinel	5	5	1	1	1	0,29

R.I. : 0,0559

IA local	Wazuh	Security Onion	Splunk	IBM QRadar	Sentinel	PESOS(W)
Wazuh	1	3	3	5	7	0,45
Security O...	1/3	1	1	3	7	0,21
Splunk	1/3	1	1	3	7	0,21
IBM QRadar	1/5	1/3	1/3	1	5	0,10
Sentinel	1/7	1/7	1/7	1/5	1	0,04

R.I. : 0,0125

Escalabilida	Wazuh	Security Onion	Splunk	IBM QRadar	Sentinel	PESOS(W)
Security O...	3	1	1/3	1	1/3	0,13
Splunk	5	3	1	3	1	0,34
IBM QRadar	3	1	1/3	1	1/3	0,13
Sentinel	5	3	1	3	1	0,34

R.I. : 0,0000

Comunidad	Wazuh	Security Onion	Splunk	IBM QRadar	Sentinel	PESOS(W)
Security O...	1	1	1/3	1/3	1/3	0,09
Splunk	3	3	1	1	1	0,09
IBM QRadar	3	3	1	1	1	0,27
Sentinel	3	3	1	1	1	0,27