# ZeroScam: Agente Experto en Ciberseguridad

## Introducción

El proyecto ZeroScam nace con la necesidad de ofrecer soluciones de prevención y asesoramiento en ciberseguridad para usuarios no expertos. En los últimos años, la digitalización ha aumentado significativamente la presencia online de individuos y negocios, lo que ha llevado a un incremento en el uso de tecnologías y aplicaciones con acceso a internet. Sin embargo, esta mayor conectividad también ha expuesto a los usuarios a diversos ciberataques y fraudes digitales.

Para abordar este problema, hemos desarrollado ZeroScam, un agente especializado en ciberseguridad basado en un modelo de lenguaje de gran tamaño (LLM). Este agente está diseñado para interactuar con los usuarios a través de Telegram y proporcionar múltiples funcionalidades de seguridad, tales como:

* Concienciación y asesoramiento en seguridad.
* Comprobación de direcciones IP maliciosas.
* Detección de páginas web fraudulentas.
* Reconocimiento óptico de caracteres (OCR) para el análisis de phishing.
* Recuperación de información especializada en ciberseguridad mediante RAG (Retrieval-Augmented Generation).
* Respuesta a consultas generales en el ámbito de la ciberseguridad.

## Equipo de Desarrollo

ZeroScam ha sido desarrollado por un equipo de expertos en inteligencia artificial y ciberseguridad:

* R. Laborda
* C. Aunión
* J. González
* H. Martí
* A. Zuluaga

## Resumen del Desarrollo

### Selección del Modelo

El agente de ciberseguridad ZeroScam se basa en un modelo LLM con arquitectura *decoder-only*. Durante la primera fase de desarrollo, se realizó una evaluación *zero-shot* para comparar el rendimiento de varios modelos, priorizando aquellos con un número moderado de parámetros para optimizar su implementación y reducir el consumo de recursos.

Como resultado de esta evaluación, se seleccionó DeepSeek, un modelo que demostró un mejor desempeño en tareas relacionadas con la ciberseguridad. Para mejorar su eficiencia y reducir los requerimientos computacionales, se implementó una versión cuantificada mediante QLoRA.

### Entrenamiento y Fine-Tuning

Para entrenar el modelo, se llevó a cabo una investigación exhaustiva con el fin de identificar fuentes de datos especializadas en ciberseguridad. Los datos se clasificaron en dos categorías principales:

1. Datos de entrenamiento: Utilizados para el ajuste fino (*fine-tuning*).
2. Datos estructurados: Incorporados en el contexto del modelo mediante RAG (*Retrieval-Augmented Generation*).

El proceso de *fine-tuning* se organizó a través de un pipeline ETL de cuatro etapas, consolidando finalmente los datos en una estructura única para la entrada al modelo. Cada observación generada contenía:

* Contexto.
* Pregunta.
* Respuesta esperada.

Para evaluar el rendimiento y la capacidad generalizadora del modelo, el conjunto de datos se dividió en entrenamiento, validación y test. Una vez finalizado el ajuste, el modelo se almacenó en Hugging Face, facilitando su integración en distintas aplicaciones.

Funcionalidades Adicionales

ZeroScam incorpora diversas tecnologías complementarias para fortalecer sus capacidades:

* Modelo OCR basado en Computer Vision: Permite extraer texto de imágenes y analizar intentos de phishing.
* Integración con VirusTotal: A través de una API, el agente puede analizar direcciones IP y páginas web sospechosas, determinando su nivel de riesgo.
* Sistema RAG con Base de Datos Vectorial: Se ha implementado una base de datos vectorial que almacena documentación técnica en ciberseguridad, permitiendo recuperar información especializada en tiempo real.

### Infraestructura y Despliegue

El agente ha sido desplegado en una instancia EC2 de Amazon, donde todos los procesos se ejecutan dentro de un entorno Docker. Para ello, se han configurado los siguientes archivos:

* Dockerfile (definición de la imagen de contenedor).
* docker-compose.yml (orquestación de servicios y dependencias).
* Archivos de configuración (.env y requirements.txt).

El código fuente se encuentra alojado en GitHub, facilitando su mantenimiento y mejora continua.

### Interfaz de Usuario

Para hacer que ZeroScam sea accesible y fácil de usar, se ha integrado con un bot de Telegram. Esta interfaz permite a los usuarios interactuar con el agente de forma rápida y eficiente, obteniendo asistencia en ciberseguridad en cualquier momento.

## Conclusión

ZeroScam representa una solución innovadora para la protección de usuarios ante ciberamenazas, combinando inteligencia artificial, visión por computadora y recuperación de información especializada. Con su integración en Telegram, ofrece un acceso sencillo y directo a herramientas avanzadas de seguridad, ayudando tanto a individuos como a empresas a mejorar su protección en el entorno digital.