

Desafío

El equipo de Data Security de Mercado Libre, se encarga de securizar las aplicaciones internas en Meli, dichas aplicaciones son utilizadas por colaboradores internos para diferentes fines. Uno de los mayores desafíos de este equipo es poder determinar si un usuario realiza acciones anómalas o indebidas sobre dichas aplicaciones y generar de forma ágil mecanismos detectivos que permitan evitar dicha actividad a tiempo.

Estamos investigando si la AI podría agilizar la creación y actualización de los mecanismos detectivos, para esto nos pidieron que desarrolláramos un sistema multiagente, el cual pueda llevar de punta a punta un análisis de vectores de ataques reales sin necesidad concreta de interacción con un analista de seguridad.

Aunque la gestión de estos hallazgos y la creación de mecanismos de seguridad en un marco corporativo es diferente, nos interesa identificar si esta tecnología es viable para realizar estas tareas de forma autónoma y luego escalarla en un flujo más complejo.

Objetivo

Realizar un sistema multiagente que posea las características mencionadas a continuación y pueda realizar las tareas propuestas para la creación de nuevos mecanismos detectivos:

Sistema: La solución debe ser una arquitectura multiagente que implemente un pipeline de análisis de vectores de ataque/riesgos, comparativa con el contexto del ecosistema a evaluar, identificación de detectores prioritarios de anomalías y generación de un reporte base para la creación y gestión de los mismos que incluya oportunidades de accionables.

- Este sistema debe tomar como dato el informe [“Data Breach Investigations Report 2025”](#) el cual será la base para comparar e identificar necesidades de implementación.

Tareas del flujo de agentes:

- **Comparativa entre contexto y datos de reporte/s:**
 - Se debe poder tomar input de usuario sobre su ecosistema, este input podrá ser un texto descriptivo sobre un sistema particular o bien una serie de documentos que describa una arquitectura/ecosistema específico a monitorear.

- El sistema deberá tomar dicha información y procesarla a fin de utilizar todo el contexto (datos del reporte + datos de input) para identificar detectores prioritarios a desarrollar.
- **Análisis de riesgos:**
 - El sistema deberá poder asociar los detectores a priorizar a algún concepto del framework “[MITRE ATT&CK techniques](#)” en el caso de que sea posible o solicitar más información al usuario sobre el contexto del sistema para lograr tal fin.
 - Mediante dicha asociación, el sistema deberá poder clasificar los riesgos de seguridad e impacto que pueden tener dichos detectores y priorizarlos para reportar.
- **Generación del reporte:**
 - Se deberá generar un reporte que no solo contenga de forma clara los detectores propuestos, el análisis de riesgos y el racional detrás de cada caso sino que también una propuesta de accionables indicando los pasos a tomar para el/los equipos de seguridad que queramos involucrar en el desarrollo de estos detectores.

Características necesarias:

- Diferentes agentes deberán tener roles y responsabilidades específicas, permitiendo interacciones dinámicas y bajo demanda. Esto significa que los agentes no operarán en un flujo secuencial, sino que podrán invocarse entre sí para optimizar la colaboración y la eficiencia en la creación final del reporte.
 - **Ejemplo 1:** El agente que se encargue de la asociación de riesgos/impacto podría necesitar llamar nuevamente al agente que se encarga de comparar con el informe de data breaches si detecta que el user le dio nuevas características de su ecosistema a tener en cuenta que no poseía a la hora de relevar los detectores prioritarios a desarrollar.
 - **Ejemplo 2:** Cualquiera de los agentes podría requerir en algún momento información o validación adicional del user para ejecutar alguno de sus pasos.
- Creación de MCP/s para agentes: Desarrollar o utilizar MCPs (Model Context Protocols) existentes para gestionar/otorgar múltiples herramientas a los agentes del sistema para las tareas necesarias de cada rol.
 - Al menos 1 MCP es requerido, se permite utilizar MCP desarrollados por la comunidad:
 - **Ej:** <https://github.com/search?q=mitre-attack-mcp&type=repositories>.
- Incluir manejo de errores para interacciones fallidas.

Entregables

- Generar registros de input-output de cada agente para poder validar los pasos intermedios en todo momento y tener trazabilidad de las interacciones/decisiones,

esto puede estar en un formato a elección pero debe generarse en un almacenamiento centralizado con cada ejecución con identificadores que permitan recrear el flujo de una “sesión”.

- Reporte de seguridad final que los agentes generen (pueden incluir varios ejemplos).
- Informe de cualquier otros supuestos, comentarios, observaciones del análisis, problemas y soluciones con los que se encontró al realizar este challenge.


Bonus ★

Estos puntos no son estrictamente necesarios pero tomaremos su implementación como un “plus” en la entrega 🦵, están ordenados por prioridad:

- Implementar un modelo local que garantice resultados similares.
- Disponibilizar la app en un frontend.
- Generar un mecanismo para validar la calidad de los datos recuperados por los métodos de RAG.
- Realizar un benchmark entre modelos, prompts o de las representaciones vectoriales (embeddings).
- Generar lógica de versiones en configuraciones/prompts utilizado por los agentes.
- Dockerizar app.

Consideraciones importantes

A tener en cuenta:

- **La solución debe ser de fácil ejecución por lo tanto deben detallarse las dependencias o ser solucionada la instalación de las mismas en otro entorno mediante su inclusión en el código. (En caso de no Dockerizar)
- Se podrán crear todas las funciones complementarias que se consideren necesarias para procesar/transformar la información.
- El “input” no sera dado para este challenge ya que sera la forma de evaluar el sistema, se espera que el candidato modele una serie de inputs a fin de probar su solución antes de la entrega, algunos ejemplos pueden ser:
 -  Ejemplos de input para candidatos

¡Gracias por tu interés en sumarte a nuestro equipo!

Cuando finalices, te pedimos que nos envíes tu resolución y luego agendaremos un espacio de 30' para que, junto con la presentación de tu análisis, puedas ampliar:

- Qué decisiones tomaste al construir tu desafío.
- Otros puntos que consideres importantes.



¡Manos a la obra!