Seminario MUICR

- **Título:** Introducción a la Ciberseguridad en Sistemas de Inteligencia Artificial
- **Fecha de impartición:** 7, 8, 14 y 15 de enero 2026
- **Horario:** 18h a 20h30
- **Lugar de impartición:** Aula máster MUICR (1G-2N17)
- **Ponente:** Jesús Friginal López, Country Manager SPAIN de la empresa SCASSI Ciberseguridad

Resumen:

Los sistemas de inteligencia artificial se enfrentan a un doble reto en materia de seguridad: por un lado, están expuestos a amenazas nuevas y emergentes derivadas de su propia naturaleza —como la manipulación de modelos, la fuga de información sensible o los ataques adversariales—, y por otro, siguen siendo vulnerables a riesgos tradicionales ya conocidos en la ciberseguridad, al basarse en infraestructuras clásicas como redes, servicios web, APIs o bases de datos. Este seminario aborda esa dualidad desde una perspectiva práctica, combinando marcos consolidados como OWASP Top 10 y MITRE ATT&CK con iniciativas específicas para IA (OWASP Top 10 for LLMs, MITRE ATLAS, Maestro). Los participantes aprenderán a identificar, analizar y mitigar vulnerabilidades tanto en los componentes tecnológicos tradicionales como en los modelos inteligentes que operan sobre ellos, utilizando herramientas como Shodan, Censys y Presidio, además de simular entornos defensivos tipo SOC. Esta formación plantea una comprensión profunda de cómo integrar la seguridad en el diseño, despliegue y gobierno de la IA, garantizando sistemas confiables, éticos y resilientes frente a las amenazas del presente y del futuro.

Sesiones:
La formación se estructura en 4 bloques:

- S1 — Frameworks generalistas de seguridad aplicados a sistemas IA: Revisaremos cómo marcos clásicos como OWASP Top 10 y MITRE ATT&CK se aplican a los riesgos en despliegues de sistemas IA, y practicaremos OSINT usando herramientas como Shodan y Censys para descubrir y analizar infraestructura pública. Construiremos un script en Python que consuma sus APIs y traduciremos los hallazgos a controles de desarrollo seguro (OWASP Secure Coding Practices).
- S2 — Defensa práctica y SOC para la seguridad de sistemas IA: Realizaremos ejercicios prácticos en la plataforma THM elegidos por los estudiantes (por ejemplo, Python tooling, logging en Windows, reglas YARA) y simularemos el flujo operativo de un SOC —desde la ingesta de logs hasta la respuesta a incidentes—, para que los asistentes comprendan cómo detectar, documentar y contener amenazas de forma coordinada.
- S3 — Gobernanza y marcos para IA: Estudiaremos marcos específicos de seguridad y gobierno de la IA, como OWASP Top 10 for LLMs, OWASP AI Exchange y MITRE ATLAS, y los traduciremos a políticas operativas: definición de roles, gestión de datos y modelos, control de versiones, validación de salidas y métricas de cumplimiento. El objetivo es desarrollar una visión sólida de la gobernanza y la responsabilidad en el ciclo de vida de la IA.
- S4 — Threat modeling para agentes de IA: Aplicaremos threat modeling con Maestro en sistemas multiagente, desarrollaremos defensas y pipelines para la detección y el enmascaramiento de información sensible (por ejemplo, con herramientas de NLP) y realizaremos ejercicios controlados de evasión y ataques adversariales para comprender cómo operan los vectores ofensivos y cómo diseñar contramedidas eficaces.