«El futuro de la defensa pasa por entender cómo engañar a una IA»

CISDE Campus Internacional para la Seguridad y la Defensa

Iván Mateos Navarro

7 de Mayo

La Inteligencia Artificial se ha convertido en una herramienta cada vez más relevante para la seguridad, la defensa y la toma de decisiones, hasta el punto de estar cambiando la forma en que se analizan los datos, se interpretan los escenarios y se actúa en entornos complejos.

Este avance abre nuevas posibilidades, aunque también plantea riesgos importantes, ya que los sistemas basados en IA pueden ayudar a detectar amenazas y mejorar la respuesta operativa, pero también pueden ser confundidos, manipulados o llevados a conclusiones erróneas. En ese terreno,

donde la tecnología se cruza con la defensa y la seguridad, desarrolla su trabajo Kallisto AI.

La compañía está especializada en soluciones avanzadas para entornos civiles y militares, con una labor centrada en la simulación, la generación de datos sintéticos, la visión por computador, el análisis de sensores y el estudio de las vulnerabilidades de los sistemas inteligentes, y especialmente los que utilizan visión por computador (Computer Vision).
Al frente de Kallisto AI se encuentra Raúl Álvarez Prieto, fundador y director de la empresa, cuya trayectoria profesional ha estado vinculada a la tecnología, las telecomunicaciones, la defensa y el emprendimiento. En los últimos años, su trabajo se ha orientado hacia uno de los ámbitos más sensibles de la innovación actual, como es el uso de sistemas de inteligencia artificial en escenarios críticos.

Combina esta labor con la formación en CISDE, donde imparte contenidos sobre los fundamentos de la Inteligencia Artificial, sus aplicaciones en defensa y seguridad, así como las tendencias que pueden marcar el futuro operativo, ético y estratégico de estas tecnologías.

Desde el Observatorio, estamos encantados de charlas con el sobre cómo la Inteligencia Artificial está cambiando la defensa y la seguridad, y en definitiva el mundo, qué usos tiene en esos ámbitos, pero también qué retos plantea y qué papel pueden jugar los sistemas autónomos en los próximos años.
Bienvenido Raúl,

Su carrera se ha desarrollado entre la tecnología, las telecomunicaciones, la defensa y el emprendimiento. ¿Qué le llevó a centrar su trabajo en la Inteligencia Artificial aplicada a la seguridad y la defensa?

Mi trayectoria siempre ha estado ligada a sistemas complejos y a entornos donde los errores no son triviales. En seguridad y defensa, una mala decisión no es un bug: tiene consecuencias reales. Cuando la Inteligencia Artificial empezó a entrar con fuerza en estos ámbitos, me interesó menos la promesa de "hacerlo todo mejor" y más una pregunta previa: ¿qué pasa cuando estos sistemas fallan o interpretan mal el entorno? Ahí vi un espacio crítico que estaba poco explorado.

Su labor docente dentro de CISDE, combina una parte conceptual con otra claramente aplicada. Desde esta experiencia dual. ¿Considera que la IA es accesible para cualquiera?

La Inteligencia Artificial es accesible para cualquiera en cuanto a herramientas y tecnología disponible, pero no lo es tanto en cuanto a comprensión real. Hoy entrenar un modelo está al alcance de mucha gente; lo difícil es entender qué asume ese modelo sobre el mundo, dónde puede fallar y por qué.

Por eso, en el Curso de IA para Defensa empezamos siempre por una visión general de qué es realmente la IA hoy en día, y sus aplicaciones concretas en el ámbito civil. A partir de ahí, pasamos a ver cómo se está utilizando actualmente en el ámbito de la defensa y la seguridad, con ejemplos reales. Y cerramos con un tercer módulo centrado en aspectos que a menudo se pasan por alto: las implicaciones éticas, legales y morales, pero también cuestiones prácticas como los recursos humanos, las herramientas de MLOps o los retos organizativos que aparecen cuando estos sistemas se llevan a producción.

Cuando se habla de IA, muchas veces se oscila entre el entusiasmo y el temor. ¿Estamos exagerando sus capacidades o seguimos sin entender la profundidad del cambio que está provocando?

Creo que hacemos ambas cosas a la vez. Exageramos lo que la IA puede hacer por sí sola, pero subestimamos el cambio estructural que introduce. No es solo

una cuestión de automatizar tareas: estamos delegando percepción, priorización y, en algunos casos, decisiones. Eso cambia profundamente cómo operan los sistemas y cómo debemos diseñarlos.

La IA ya forma parte de muchas actividades civiles, está presente en la medicina, la industria o las comunicaciones. ¿Qué cambia cuando esas mismas tecnologías se trasladan al ámbito de la defensa y la seguridad?

Cambia el nivel de exigencia. En defensa no basta con que algo funcione "la mayoría del tiempo". Hay adversarios, hay intención de engaño y hay consecuencias operativas. Un sistema de IA en defensa debe enfrentarse a entornos que no cooperan y eso obliga a pensar más en robustez que en pura precisión.

Kallisto AI nace como una empresa tecnológica especializada en la aplicación práctica de la IA para entornos civiles y de defensa. ¿Qué necesidad concreta detectaron para poner en marcha la compañía?

Detectamos que se hablaba mucho de entrenar y desplegar sistemas de IA, pero muy poco de **qué ocurre cuando esos sistemas operan en entornos que no son benignos**.

La mayoría de los desarrollos asumían, de forma implícita, que el entorno físico iba a cooperar con el algoritmo, y eso en defensa rara vez es cierto.

En ese contexto vimos un nicho muy claro en **camuflaje, ocultación y engaño frente a algoritmos de IA**, un ámbito en el que prácticamente nadie estaba trabajando de forma comercial. Había doctrina y experiencia histórica orientada al ser humano, pero muy poco pensado para sistemas de percepción automática. Por eso decidimos centrarnos en simulación avanzada, generación de datos y análisis de percepción, para entender dónde y por qué estos sistemas pueden equivocarse cuando el entorno está diseñado activamente para confundirlos.

Muchas empresas hablan hoy de Inteligencia Artificial, pero pocas se especializan en cómo pueden fallar o ser engañados estos sistemas. ¿Por qué decidieron situarse precisamente en ese terreno?

Porque en defensa y seguridad no basta con saber cuándo un sistema acierta; hay que saber cuándo y cómo se equivoca. Muchos sistemas funcionan muy bien en pruebas controladas, pero fallan de maneras inesperadas cuando las condiciones cambian. Eso no es un detalle técnico, es una vulnerabilidad.

En defensa, una tecnología no se valora solamente por ser innovadora, sino también por su utilidad real en operaciones. ¿Qué necesidad concreta de las fuerzas armadas o de los operadores de seguridad viene a cubrir Kallisto?

Desde 2022 teníamos bastante claro que la IA iba a cambiar muchas cosas, y entre ellas, la forma en la que se desarrollarían los conflictos. Sin entrar en escenarios extremos o de ciencia ficción, era evidente que el aumento de satélites, drones y medios de vigilancia, junto con la reducción drástica de sus costes, iba a hacer que el campo de batalla fuese cada vez más visible. Lo que algunos llaman una reducción de la niebla de la guerra también implica que detectar movimientos de tropas o activos sería mucho más sencillo.

Antes de que eso se convirtiera en una realidad plenamente operativa, queríamos hacer algo por nuestros soldados. Kallisto Shield nace precisamente de ahí y de una observación muy concreta: la mayoría de los sistemas de visión artificial asumen que el mundo físico es estable y fácilmente interpretable. Nosotros no buscamos engañar a un modelo concreto ni "cegar" a la IA, sino actuar sobre la capa de percepción, introduciendo ambigüedad estructural en lo que estos sistemas observan para reducir su fiabilidad de forma persistente. Es una manera de devolver complejidad al entorno y proteger a quienes operan en él.

Su empresa se define como una empresa *deep-tech*. ¿Qué significa eso en la práctica?

En la práctica significa que no vendemos resultados inmediatos ni soluciones superficiales. Trabajamos a nivel de fundamentos: percepción, sensores, datos y modelos, pero también en entender cómo funcionan realmente los sistemas que hoy se están desplegando.

Conocemos bastante bien cómo operan los algoritmos de IA actuales, pero también cómo funcionan las cámaras, los sensores, los satélites, los drones y las plataformas que los integran. Todo eso condiciona qué ve un sistema, cómo lo interpreta y dónde puede fallar. Por eso diseñamos nuestras soluciones teniendo en cuenta todo el conjunto, no solo el algoritmo aislado. Es un trabajo más lento, pero necesario si se quiere construir tecnología que funcione en el mundo real y no solo en el laboratorio.

Uno de los servicios pasa por generar escenarios simulados con distintas condiciones de terreno, clima, sensores, ángulos o resoluciones. ¿Por qué esa capacidad puede ser tan valiosa para entrenar y evaluar sistemas antes de llevarlos al mundo real?

Porque muchos escenarios relevantes son difíciles, caros o imposibles de reproducir en el mundo real. La simulación nos permite explorar ángulos, resoluciones, sensores y condiciones límite antes de que un sistema se despliegue. Es una forma de fallar barato y aprender antes.

La generación de datos sintéticos es otra de las líneas de trabajo. ¿En qué se traduce esto cuando no existen suficientes datos reales, cuando obtenerlos es demasiado caro o cuando ciertos escenarios son difíciles de reproducir?

Permite entrenar y evaluar sistemas cuando los datos reales no existen, son insuficientes o están sesgados. Pero, más importante aún, permite diseñar

escenarios adversariales de forma controlada, algo fundamental para entender la robustez real de un sistema.

Muchos avances tecnológicos fracasan no por la idea, sino por la integración. ¿Qué importancia tiene diseñar soluciones que puedan encajar con sistemas ya existentes y no obliguen al cliente a empezar desde cero?

Es absolutamente fundamental. En entornos operativos nadie puede permitirse reemplazar vehículos, sensores o plataformas enteras cada pocos años solo para introducir una nueva tecnología. Si una solución no encaja con lo que ya existe, simplemente no se adopta.

Por eso Kallisto Shield se ha diseñado desde el inicio como un sistema **versátil y fácilmente integrable**: un kit que puede añadirse a prácticamente cualquier vehículo sin modificarlo, sin consumo energético, sin electrónica activa y con un coste contenido. Es una solución pasiva, robusta y pensada para el uso real, no para un laboratorio. Nuestro objetivo es que funcione como **una capa más**, que se pueda desplegar rápido y a escala, y que aporte valor sin cambiar la forma habitual de operar de los usuarios.

En el ámbito de la defensa se habla mucho de rapidez, pero también de seguridad, validación y responsabilidad. ¿Cómo se equilibra la necesidad de innovar rápido con la obligación de no desplegar tecnologías inmaduras?

Aceptando que no todo lo que es técnicamente posible debe desplegarse de inmediato. Innovar rápido no significa hacerlo a ciegas. Hay que probar, validar, equivocarse y entender antes de llevar una tecnología al entorno operativo, y ser honesto con los límites de lo que se está desarrollando.

Por eso damos mucha importancia a herramientas como los gemelos digitales, la simulación del campo de batalla y la generación de datos sintéticos. Nos permiten explorar escenarios complejos, probar comportamientos límite y detectar fallos sin poner en riesgo a personas ni sistemas reales. Además, este enfoque reduce de forma significativa costes y tiempos de desarrollo, algo

especialmente importante para una start-up, donde la eficiencia y la buena gestión del flujo de caja son clave. De esta forma podemos avanzar rápido, pero con control y responsabilidad.

Uno de los conceptos más interesantes de su trabajo es el engaño aplicado a los algoritmos de IA y lo cierto es que el uso de señuelos no es nuevo en la historia militar, pero ahora el observador puede ser un algoritmo. ¿Qué cambia cuando el engaño no va dirigido al ojo humano, sino a un sistema de visión artificial?

Cambia prácticamente todo. Un algoritmo no "sospecha" ni interpreta el contexto como lo hace una persona; **confía en patrones estadísticos y en la coherencia de los datos que recibe**. Eso significa que el engaño ya no puede basarse solo en lo que parece creíble para un ser humano, sino en cómo se construye la percepción de una máquina.

Con Kallisto Shield no buscamos engañar un clasificador concreto ni atacar un modelo específico. Diseñamos el sistema con un conocimiento profundo de cómo funcionan hoy los algoritmos de visión artificial, pero también de cómo capturan información los sensores: cámaras visibles, infrarrojas, multiespectrales e incluso radar SAR. El Shield actúa sobre el mundo físico de forma que **afecta simultáneamente a todas esas bandas**, incluidos dominios donde tradicionalmente el camuflaje era mucho más complejo.

Esto permite, por un lado, **proteger vehículos reales** reduciendo la fiabilidad de su detección e identificación, y por otro **generar señuelos muy baratos**, pasivos y fácilmente desplegables. Lo relevante no es un señuelo concreto, sino que el sistema permite **generar millones de firmas electromagnéticas distintas**, todas físicamente plausibles, lo que complica enormemente el aprendizaje y la discriminación por parte de los modelos de IA.

En la práctica, lo que hacemos es **ecualizar los datos de entrada** que alimentan a estos sistemas de percepción automática: vehículos reales y señuelos

empiezan a parecerse demasiado entre sí desde múltiples ángulos, resoluciones y sensores. Cuando ocurre eso, la suposición de que el entorno es estable y separable deja de cumplirse, y métricas clásicas como precisión o recall empiezan a fallar de manera conjunta. La IA sigue viendo cosas, pero **deja de verlas con fiabilidad**, y eso tiene implicaciones operativas muy profundas.

Conflictos y escenarios como el de Ucrania han acelerado la preocupación por los drones y por los sistemas autónomos de ataque. ¿Hasta qué punto ese escenario ha reforzado la necesidad de soluciones como las que desarrolla su compañía?

Conflictos como el de Ucrania han dejado claro que los drones y los sistemas autónomos no son un escenario futuro, sino una realidad plenamente operativa. También han evidenciado que **la percepción se ha convertido en un nuevo campo de batalla**: quien controla cómo se observa y se interpreta el entorno tiene una ventaja decisiva.

La proliferación de sensores, drones y medios de vigilancia de bajo coste ha hecho que los movimientos sean cada vez más visibles y rastreables. En ese contexto, recuperar cierta complejidad y ambigüedad en el campo de batalla ya no es solo deseable, sino necesario. Este tipo de escenarios refuerzan claramente la necesidad de soluciones que no solo detecten mejor, sino que **limiten la fiabilidad de la percepción automática del adversario**.

En Europa se habla cada vez más de soberanía tecnológica en defensa. ¿Qué puede aportar una empresa española como Kallisto en un mercado donde muchas capacidades críticas tradicionalmente vienen de fuera?

Puede aportar soberanía técnica, pero también algo igual de importante: pensamiento crítico y capacidad de cuestionar lo establecido. Las startups tenemos un margen de libertad y una presión por innovar que muchas

empresas más consolidadas, con productos maduros y contratos en curso, difícilmente pueden permitirse.

En Kallisto tenemos esa "hambre" por explorar campos nuevos, asumir riesgos técnicos y meternos en problemas complejos que todavía no tienen solución clara. En ese sentido somos, como se diría en inglés, un troublemaker: removemos el avispero, cuestionamos suposiciones que se dan por buenas y abrimos líneas de trabajo que no encajan fácilmente en catálogos existentes. Pero lo hacemos con un objetivo muy claro: contribuir al bien común y a que Europa desarrolle criterio propio sobre cómo diseñar, usar y proteger sus sistemas tecnológicos.

No se trata de replicar lo que ya existe fuera, sino de pensar diferente y aportar capacidades nuevas allí donde otros no están mirando. Ahí es donde creemos que una empresa española, pequeña pero muy especializada, puede marcar la diferencia.

En el debate público, la defensa tecnológica y especialmente el uso de la IA a veces se mira con recelo. ¿Cómo tranquilizaría a los más escépticos?

Entiendo ese recelo, pero creo que es importante afrontar el debate con realismo. Todas las grandes —y también muchas pequeñas— potencias ya están investigando, desarrollando y desplegando sistemas autónomos y tecnologías basadas en IA. Ese proceso está en marcha y no va a detenerse. En ese contexto, no hacer nada o mirar hacia otro lado no es una postura ética: sería una irresponsabilidad.

Nuestro trabajo no busca automatizar la guerra ni eliminar el factor humano, sino entender mejor los límites de la automatización y cómo se comportan estos sistemas en entornos reales y adversos. La IA puede ayudar a proteger vidas si se usa con conocimiento, control y cautela. Pero ignorar sus vulnerabilidades, asumir que siempre funcionará bien o dejar que otros

marquen las reglas sin reflexión propia sería mucho más peligroso que abordarlo con espíritu crítico y responsabilidad.

Más allá de Kallisto Shield, ¿En qué otros proyectos trabaja la compañía?

Además de Kallisto Shield, en Kallisto AI trabajamos en simulación multisensor, evaluación de sistemas autónomos, generación de escenarios complejos y estudios de robustez tanto para clientes civiles como de defensa. Todo ello con un enfoque muy práctico y orientado a entender cómo se comportan estos sistemas fuera de entornos ideales.

En esa línea hemos creado también Kallisto Deception Lab, con la ambición clara de convertirnos en un referente mundial en AI deception. Es un espacio donde investigamos, probamos y desarrollamos técnicas de engaño dirigidas específicamente a sistemas inteligentes, combinando simulación avanzada, datos sintéticos y análisis profundo de percepción automática. Nuestro objetivo es anticiparnos a cómo evolucionan estos sistemas y entender, antes que nadie, dónde y por qué pueden ser engañados, para diseñar soluciones que funcionen en escenarios reales y futuros.

La aplicación de IA en defensa puede mejorar la protección, la anticipación y la respuesta, pero también plantea riesgos evidentes. ¿Dónde cree que debe estar la línea que no debería cruzarse?

En delegar decisiones irreversibles sin comprensión ni control humanos adecuados. La IA puede asistir, priorizar y alertar, pero **la responsabilidad última no puede desaparecer tras un algoritmo**.

Y, para terminar, si tuviera que resumir el reto de los próximos años, ¿diría que se trata de aprender a usar mejor estas herramientas, de aprender a confiar en ellas o de aprender a desconfiar de ellas con criterio?

Aprender a **desconfiar con criterio**. No se trata de rechazar la IA ni de confiar ciegamente en ella, sino de entender cuándo funciona, cuándo no y por qué. Solo así podremos usarla de forma responsable en entornos críticos.