Asamblea PLANETIC

Juan – Román Martinez Arranz

06/11/2024

Conceptos

Frugalidad

- Capacidad de poder generar Modelos que funcionen perfectamente a pesar de tener escasez de datos

Robustez

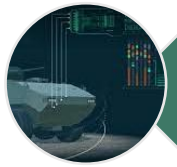
- Capacidad de un Modelo de IA para seguir funcionando correctamente a pesar de tener algunos juegos de datos “anómalos”, circunstancias nuevas/cambiantes y/o ataques enemigos

Explicabilidad

- Capacidad del sistema para explicar un Modelo de IA y sus razonamientos seguidos para llegar a proponer una solución o acción o decisión.

Human-in-the-loop:

- Integrar al ser humano en la toma de decisiones críticas.



Gemelo digital de IA:

- Copia digital de un sistema para replicar comportamientos, en este caso, sistemas de IA

Fusión de datos Inteligente:

- Recolección de datos de diferentes fuentes y tipos que se combinan y usan en Modelos de IA para descubrimiento de nuevas amenazas y mejorar la consciencia situacional

Nuestra perspectiva

La **escasez** de datos (frugalidad) es muy común en los sistemas de Defensa

La **Robustez** cobra una vital importancia para evitar que, ya sea por error del operario o por un ataque malicioso, los Modelos de IA empiecen a fallar

La **explicabilidad**, es un habilitador que permita detectar cuando un Modelo de IA esta dando resultados no esperados, donde **el human-in-the-loop** es una pieza clave.

Todo sistema que emplee IA debe, desde los estadios iniciales, contemplar los mecanismos para asegurar la **robustez y explicabilidad**

Esquema análogo y complementario **al security-by-design, SecDevOps**, y complementarse con mecanismos de **Zero-Trust**

Los elementos que se intuya que pueden estar comprometidos y/o no funcionando como se espera, deben poder ser aislados del sistema global de forma rápida y eficiente, permitiendo la continuidad de las operaciones ➔ **Ciber-resiliencia**

Nuestra perspectiva

Analizamos **confiabilidad y ciberseguridad de la IA** para equiparla con mecanismos confiables lo que requiere analizar de manera iterativa todos los aspectos que pueden influir en el rendimiento de los modelos de IA (desde la recopilación de datos hasta la implementación del modelo en las aplicaciones finales)

Estas consideraciones son aún más ciertas cuando consideramos la capacidad de los sistemas de inteligencia artificial para resistir y contrarrestar los **ciberataques** y las **amenazas híbridas**.

La **confiabilidad y ciberseguridad** debe planificarse desde la fase de diseño para evitar volver a planificar el proceso de construcción de IA, retrasando el despliegue de los modelos. Especialmente crítico en situaciones de ciberseguridad que requieren la implementación de contramedidas proactivas en períodos de tiempo razonables.

Ls modelos de IA pueden ser atacados (intencionalmente o por error del operador) para perturbar su proceso de inferencia, lo que puede causar **consecuencias críticas** para la vida de las personas, la sociedad y los sistemas de Defensa. El foco estará en el Impacto.

Nuestro enfoque acelerar el análisis de ciberataques a sistemas a través de aprendizaje automático utilizando **gemelos digitales e IA generativa.**

Ciber consciencia situacional y Toma decisiones

EU GUARDIAN

ECYSAP
European Cyber Situational Awareness Platform

Automatización

- **Operación automatizada** de procesos de gestión de incidentes mediante Inteligencia Artificial.
- Foco: Procesos de **gestión de incidentes y ciberdefensa**.
- Detección, mitigación y respuesta a los retos de seguridad de forma semiautomática o automática.
- Apoyo a operadores humanos, analistas: **Toma de decisiones**.
- **Robustecimiento** de las infraestructuras militares y la protección contra las ciber amenazas avanzadas.

Consciencia Situacional

Desarrollo e implementación de fundamentos teóricos innovadores, métodos, prototipos de investigación y la integración hacia la provisión de una **plataforma operacional Europea** para la gestión de la **ciber conciencia situacional en tiempo** real con capacidades de **respuesta defensiva** rápida y apoyo a la toma de decisión de los usuarios finales militares.

Frugalidad, Robustez y Explicabilidad

FaRADAI
Frugal Robust Advanced Intelligence

Frugalidad

Desarrollo de algoritmos para **ATD/ATR** mediante procesado de imágenes.
Debido al escaso número de imágenes disponibles, se entrenan los modelos para que sean capaces de **detectar y reconocer objetivos** a pesar de ello, consiguiendo resultados precisos

Escalabilidad e Interoperabilidad

Se están desarrollando algoritmos **modulares, interoperables y escalables** para poder ampliarse su uso y funcionalidades
Integración entre sistemas ATD/ATR tierra-tierra y aire-tierra para mejorar la consciencia situacional

Robustez y explicabilidad

Los modelos deben son capaces de dar resultados precisos bajo **“deceptive conditions”**, tales como niebla, falta de luz o LoS perturbada por orografía. Además los modelos desarrollados mantienen la precisión en caso de recibir datos manipulados o erróneos. Todas las decisiones sugeridas se (auto) explican de forma eficiente a los operadores y usuarios para que puedan valorar su posible uso, así como ayudar a detectar cuando los algoritmos podría estar fallando.

MLOPS para seguridad y vigilancia submarina

SMAUG

Vigilancia Marítima y Submarina

Trabajamos en potenciar la seguridad de los puertos y fronteras marítimas, mejorando la vigilancia y seguridad, incluyendo la **submarina** aprovechando el poder de la Inteligencia Artificial (IA), utilizando un sistemas integrados capaz de proporcionar datos sobre detección y análisis de amenazas entre 3 elementos principales:

- Infraestructura de seguridad portuaria
- Sistemas avanzados de detección **submarina**
- Embarcaciones de vigilancia con capacidades mejoradas

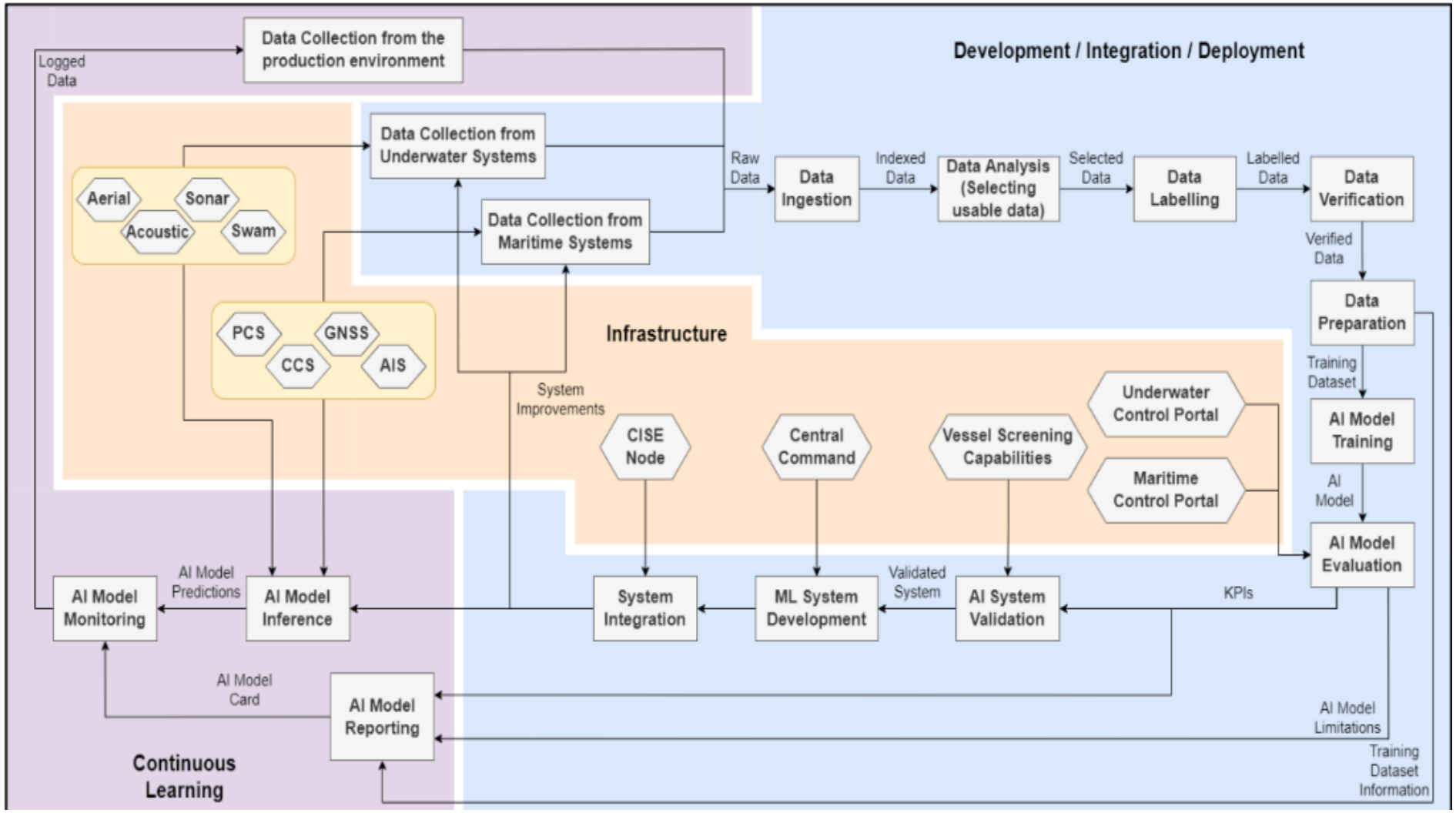
Sistemas

- Detección acústica (hidrófonos)
- Sonar para de exploración rápida de casco de embarcaciones
- Inspección por sonar de alta resolución
- Ubicación autónoma colectiva enjambre de UUV

Fuentes de datos: PCS, CCS, AIS, CISE, GNSS

- **UC1**: Detección de Narco-Submarinos y explosivos. (Valencia)
- **UC2**: Detección de tráfico y transporte ilegales (Elefsina y Heraklion)
- **UC3**: Prevención y detección de sabotajes, tipo Nordstream (Drammen)

MLOPS para seguridad y vigilancia submarina



Gemelos digitales y Modelos IA

ARMOR

Reinforcement Learning

Un sistema que recopile toda la información del **gemelo digital** puede ayudar al supervisor humano a tomar decisiones proactivas y a ajustar la solidez de los modelos de IA aplicando recomendaciones al proceso de aprendizaje automático original.

Seguridad

Al mismo tiempo, el acceso y el control generales de la arquitectura subyacente de la IA deben ser seguros con **mecanismos sólidos** que protegen el cálculo y el almacenamiento

Desarrollo y testing con Gemelos Digitales

La idea principal es que los gemelos digitales se creen a partir de aplicaciones o sistemas que ejecutan modelos de IA. Este **entorno digital combinado** con técnicas generativas se puede utilizar para probar, validar y verificar exhaustivamente la solidez de los modelos ante una diversidad de ataques y contramedidas

Fusión de Datos

Ingeniería del Dato

Trabajamos con **datos heterogéneos** de forma colaborativa incluyendo la capa de gestión que maximiza el uso de los recursos.
Se trabajan las principales capacidades de fusión de datos, así como las necesidades de comunicación entre los diferentes sistemas involucrados

Toma de Decisiones

Uso de IA para mejorar el apoyo a la toma de decisiones, la **detección encubierta** y para ayudar a maximizar las capacidades encubiertas en apoyo de la Gestión Dinámica

Fuentes de datos y colaboración

El conjunto de sistemas disponibles no se debe limitar a aquellos que se encuentran en una plataforma determinada, pudiendo nutrirse de otras fuentes
El sistema permite la generación de **imágenes de referencia sintéticas** y la detección de anomalías para complementar los enfoques basados en IA/ML.


indra
At the core