COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

ASOCIACIÓN NACIONAL
DE INGENIEROS DEL ICAI

comillas.edu

Jornada Técnica.
Aplicación Nuevas tecnologías de la IA al Ferrocarril

Ciberseguridad

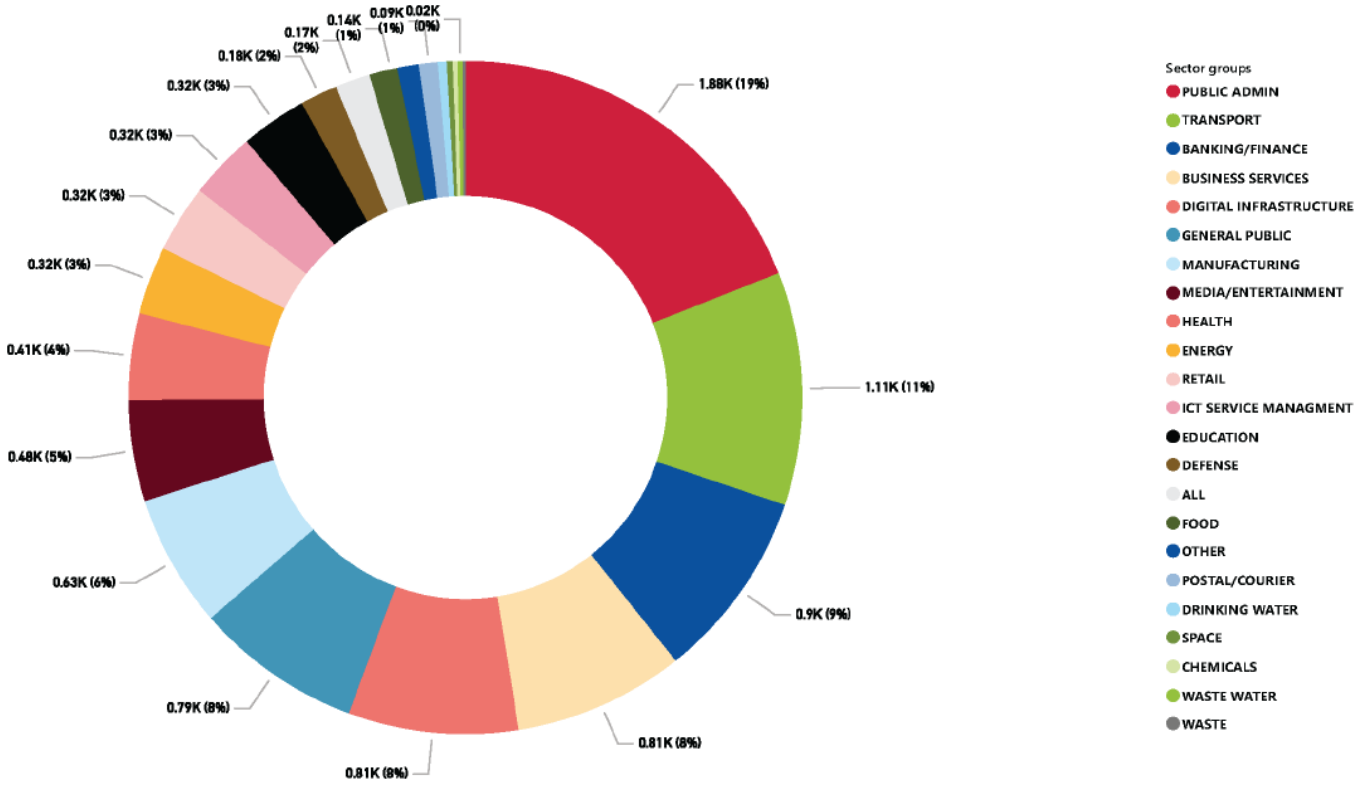
Desde la perspectiva de los sistemas IT & OT

Desde la perspectiva de la IA

Informe ENISA Threat Landscape 2024

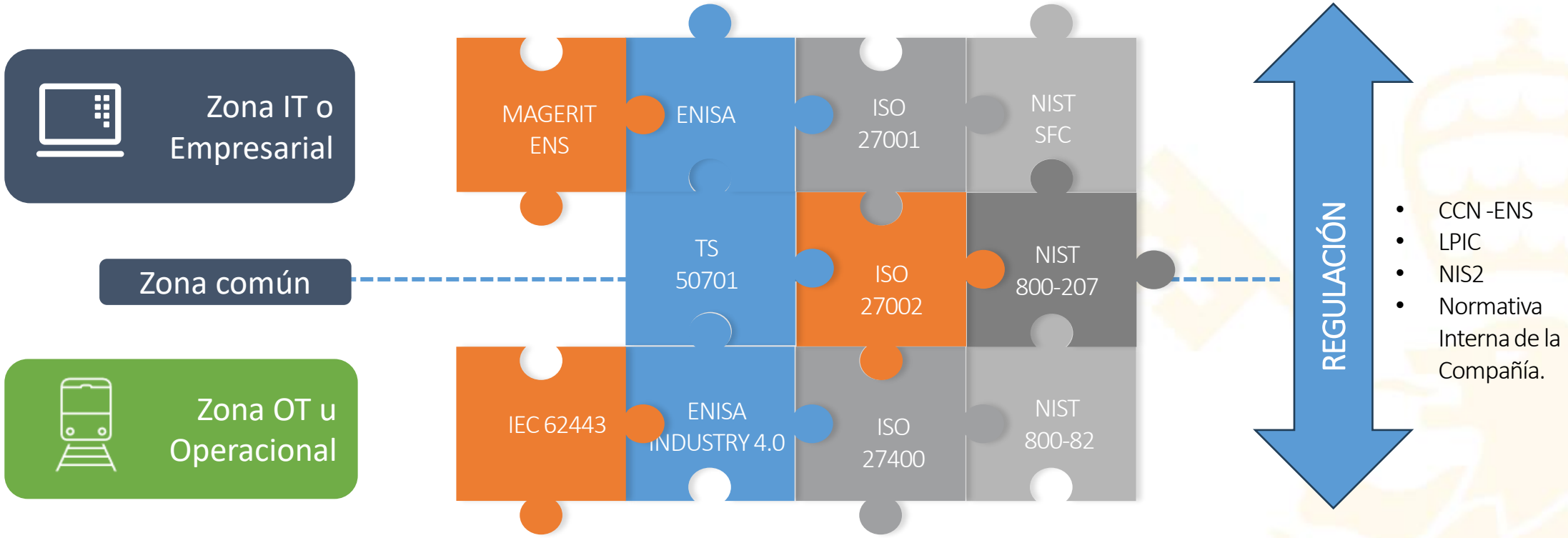
- Transportes, el segundo sector con más ciberataques (11%)

Figure 6 Targeted sectors per number of incidents (July 2023 - June 2024)

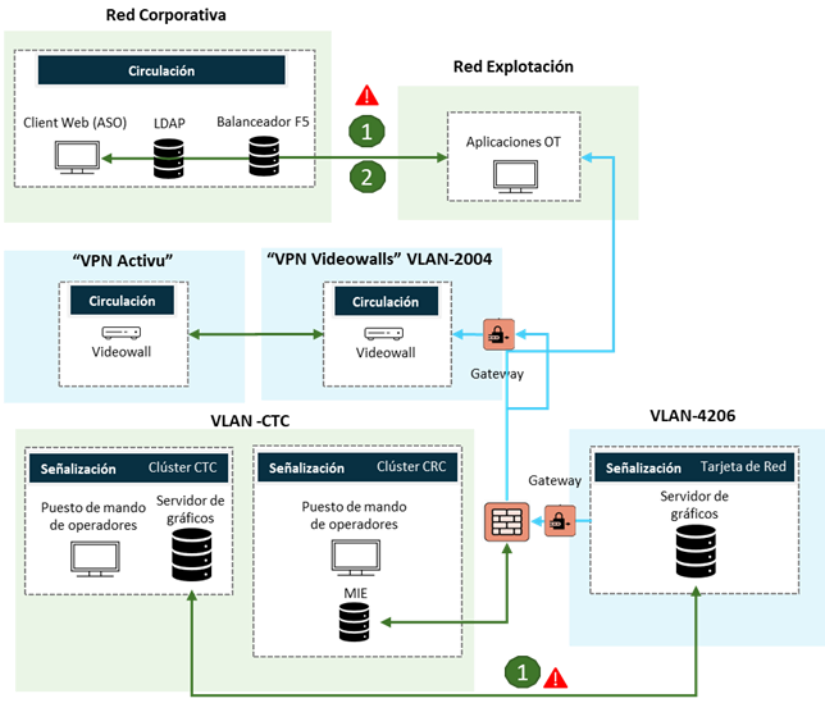


Integración y separación sistemas IT-Informáticos y OT- Operacionales

- El cumplimiento de estándares y buenas prácticas IT & OT en ciberseguridad



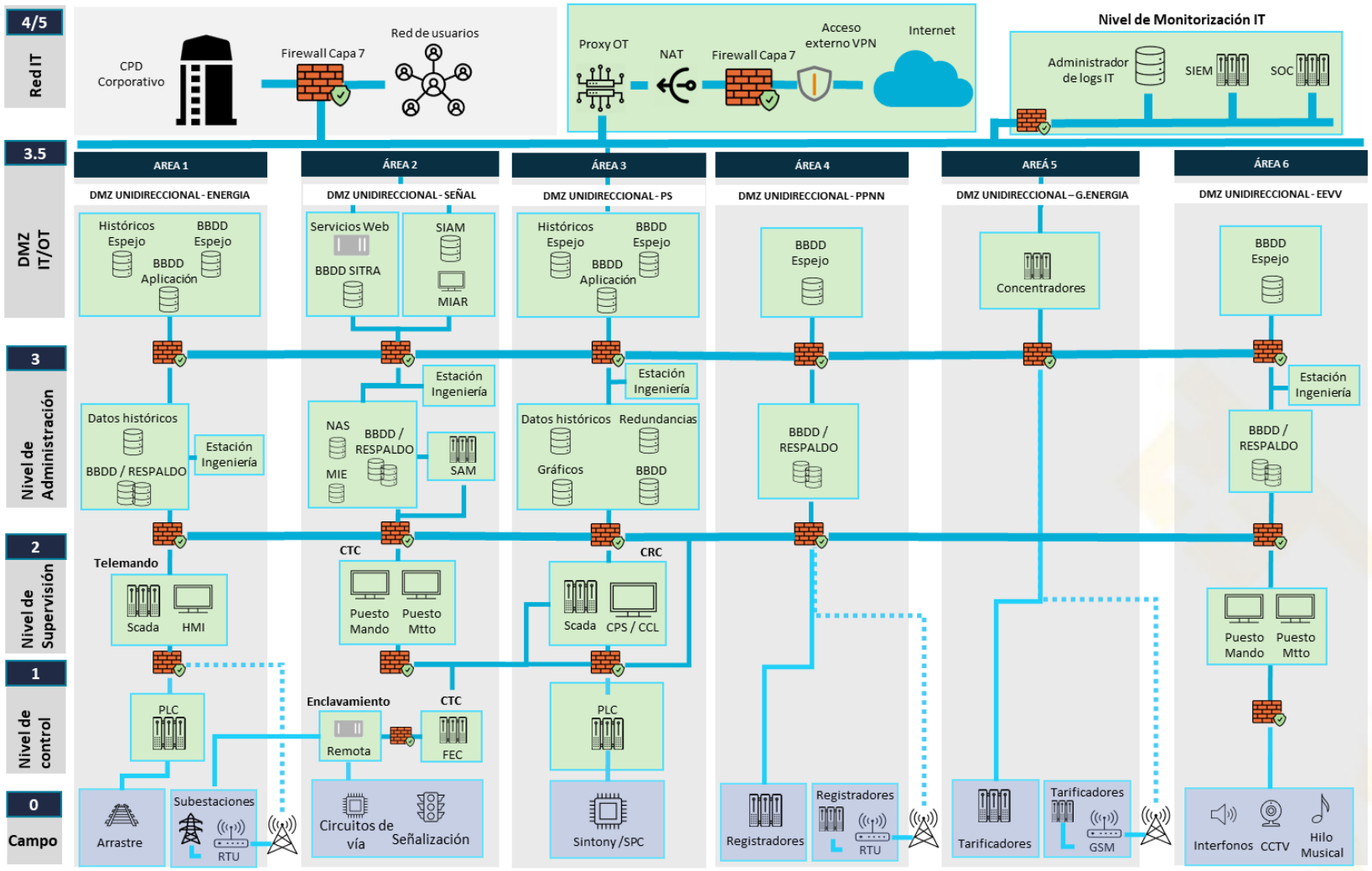
OT - Riesgos de una arquitectura tradicional



Riesgos

	Vulnerabilidad	Amenaza	Impacto
1	Autenticación insegura en dispositivos OT	Escalada de privilegios mediante falsas autenticaciones	Perdida de la protección
2	Protocolo de escritório remoto (RDP) habilitado	Captura y explotación de datos remotos	Perdida y degradación de los servicios
3	Protocolo SMBv3 utilizado en dispositivos OT basados en Windows	Ejecución código arbitrario de forma remota.	Pérdida del control de procesos
4	Versiones del software obsoleto	Escalada de privilegios y obtener acceso no autorizado	Degradación de los servicios
5	Uso de puertos inseguros	Interceptación de información para la explotación de servicios	Perdida de la disponibilidad
6	Falta de parches y actualizaciones	Explotación vulnerabilidades conocidas.	Perdida y degradación de los servicios

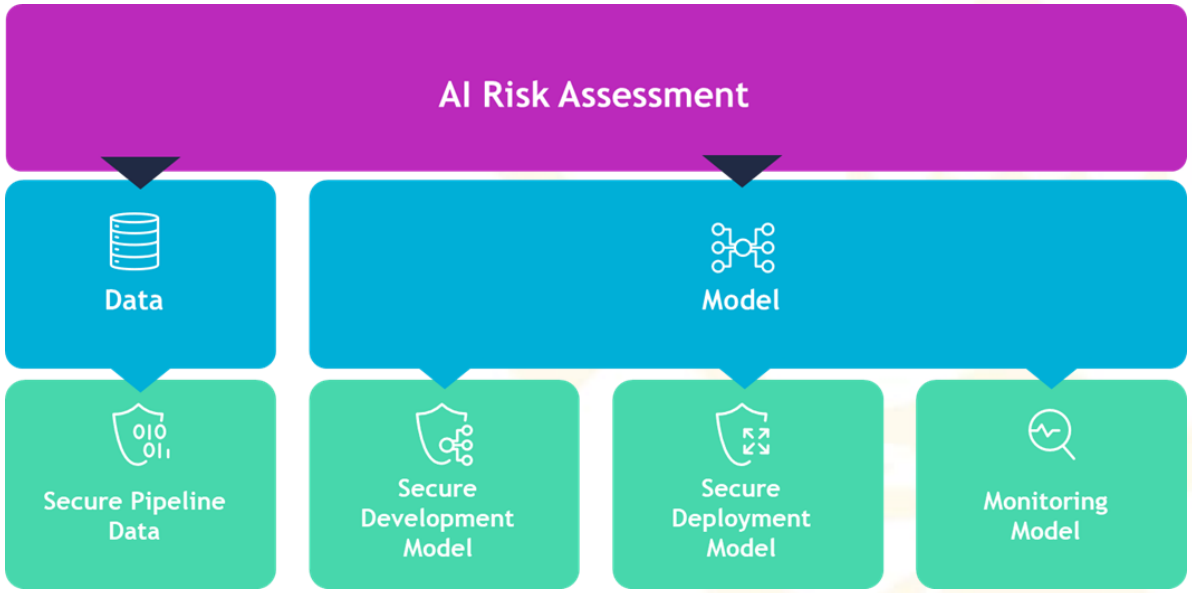
OT - Hacia una arquitectura segura - Modelo Purdue – Zero Trust



IA introduce nuevos elementos a proteger…

Nuevas Amenazas - Vectores de Ataque – Vulnerabilidades - Riesgos

- Fugas y envenenamiento de Datos
- Vulnerabilidades en parametrización
- Envenenamiento de los Modelos
- Ataques en la cadena de suministro
- Vulnerabilidades en la infraestructura
- Sabotaje del comportamiento
- Robo de los modelos
- Fugas de información
- …



IA introduce nuevos elementos a proteger…

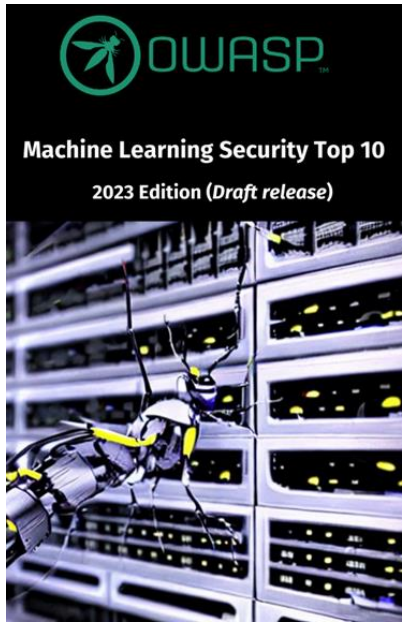
NIST – MITRE – OWASP

Los principales organismos de ciberseguridad ya disponen de modelos de riesgos – ataques - vulnerabilidades

MITRE ATLAS
Navigate threats to AI systems through real-world insights

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY U.S. DEPARTMENT OF COMMERCE
Artificial Intelligence Risk Management Framework (AI RMF 1.0)

OWASP Top 10 for LLM Applications
VERSION 1.0.1
HTTPS://LLMTOP10.COM



Top 10 Machine Learning Security Risks

- ML01:2023 Input Manipulation Attack
- ML02:2023 Data Poisoning Attack
- ML03:2023 Model Inversion Attack
- ML04:2023 Membership Inference Attack
- ML05:2023 Model Theft
- ML06:2023 AI Supply Chain Attacks
- ML07:2023 Transfer Learning Attack
- ML08:2023 Model Skewing
- ML09:2023 Output Integrity Attack
- ML10:2023 Model Poisoning

Técnicas y tácticas de ciberataques a sistemas de IA

- ATLAS - Adversarial Threat Landscape for Artificial-Intelligence Systems

MITRE ATLAS — Matrix | Tactics | Techniques | Mitigations | Case Studies | Resources

The ATLAS Matrix below shows the progression of tactics used in attacks as columns from left to right, with ML techniques belonging to each tactic below. & indicates an adaption from ATT&CK. Click on the blue links to learn more about each item, or search and view ATLAS tactics and techniques using the links at the top navigation bar. View the ATLAS matrix highlighted alongside ATT&CK Enterprise techniques on the ATLAS Navigator.

Reconnaissance & (5 techniques)	Resource Development & (9 techniques)	Initial Access & (6 techniques)	ML Model Access (4 techniques)	Execution & (3 techniques)	Persistence & (4 techniques)	Privilege Escalation & (3 techniques)	Defense Evasion & (3 techniques)	Credential Access & (1 technique)	Discovery & (6 techniques)	Collection & (3 techniques)	ML Attack Staging (4 techniques)	Exfiltration & (4 techniques)	Impact & (7 techniques)
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	AI Model Inference API Access	User Execution &	Poison Training Data	LLM Prompt Injection	Evade ML Model	Unsecured Credentials &	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model	Exfiltration via ML Inference API	Evade ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities &	Valid Accounts &	ML-Enabled Product or Service	Command and Scripting Interpreter &	Backdoor ML Model	LLM Plugin Compromise	LLM Prompt Injection		Discover ML Model Family	Data from Information Repositories &	Backdoor ML Model	Exfiltration via Cyber Means	Denial of ML Service
Search Victim-Owned Websites	Develop Capabilities &	Evade ML Model	Physical Environment Access	LLM Plugin Compromise	LLM Prompt Injection	LLM Jailbreak	LLM Jailbreak		Discover ML Artifacts	Data from Local System &	Verify Attack	LLM Meta Prompt Extraction	Spamming ML System with Chaff Data
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application &	Full ML Model Access		LLM Prompt Self-Replication				LLM Meta Prompt Extraction		Craft Adversarial Data	LLM Data Leakage	Erode ML Model Integrity
Active Scanning &	Publish Poisoned Datasets	LLM Prompt Injection							Discover LLM Hallucinations				Cost Harvesting
	Poison Training Data	Phishing &							Discover AI Model Outputs				External Harms
	Establish Accounts &												Erode Dataset Integrity
	Publish Poisoned Models												
	Publish Hallucinated Entities												

MLOps = DevOps + ML + Cyber

Negocio - IA – Legal – Cyber

Informe ENISA Threat Landscape 2024

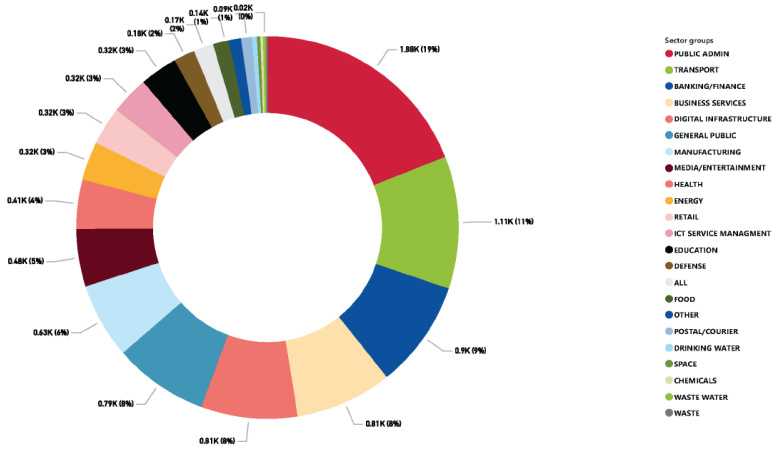
- Transportes, el segundo sector con más ciberataques (11%)

El primero es administración pública.

ADIF es administración y transporte, además de infraestructura crítica.

El nivel de protección de nuestros sistemas tiene que cumplir con los estandares más rigurosos.

Figure 6 Targeted sectors per number of incidents (July 2023 - June 2024)

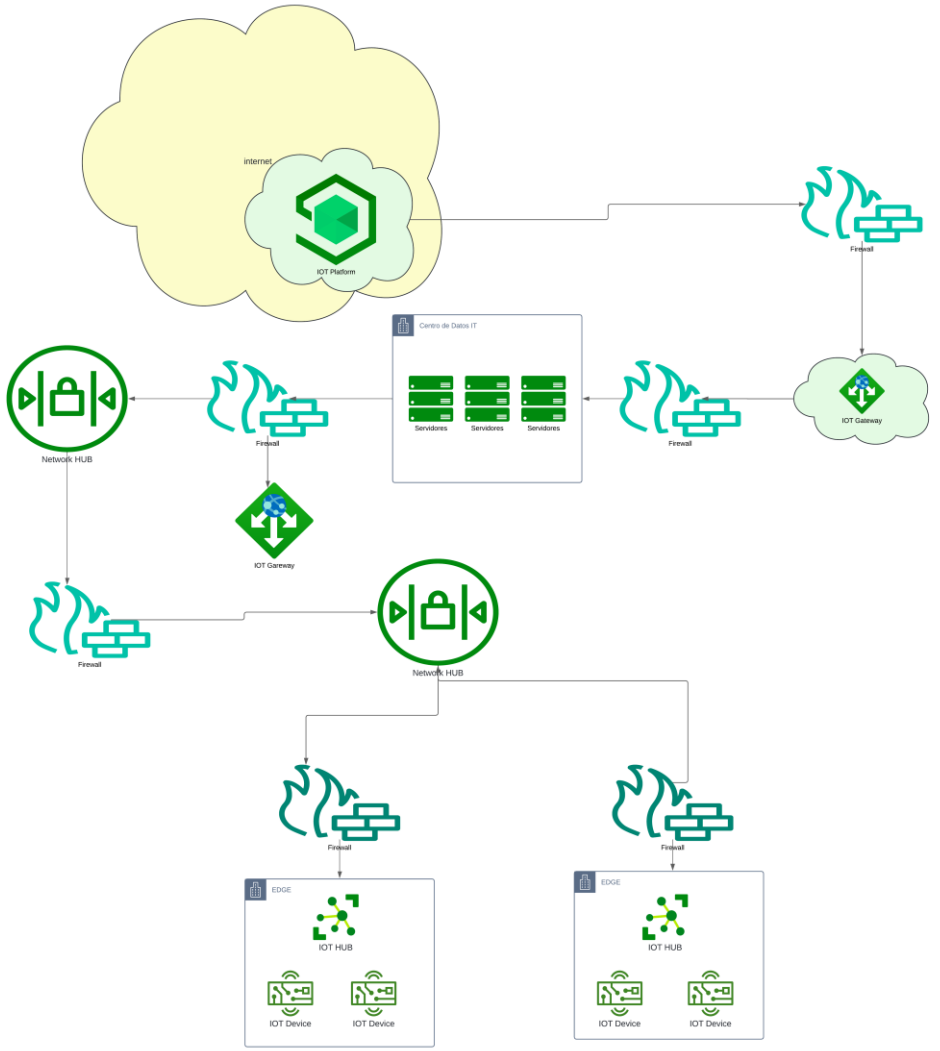


Arquitecturas Zero Trust. Proyectos IIoT

Los sistemas generan un volumen gigantesco de información.

La IA nos ayuda a poder procesarlo en tiempo cuasi real y tomar acciones.

La seguridad vs la operación no es fácil de manejar.



El trabajador aumentado

Asistentes virtuales con IA que asisten en el proceso de las operaciones ferroviarias.

Guardarrailes necesarios.
Gobierno de la IA.

Gracias

Javier Jarauta Sánchez

Director Master en Ciberseguridad Comillas ICAI

UNIVERSIDAD PONTIFICIA COMILLAS

Jesús Salvador Rueda

Jefatura de Área de Sistemas TIC

DIRECCIÓN DE TRANSFORMACIÓN DIGITAL Y SISTEMAS

ADIF