# KI-drevet sikkerhetsorkestrering, automatisering og respons for digitale tvilling baserte kritiske systemer

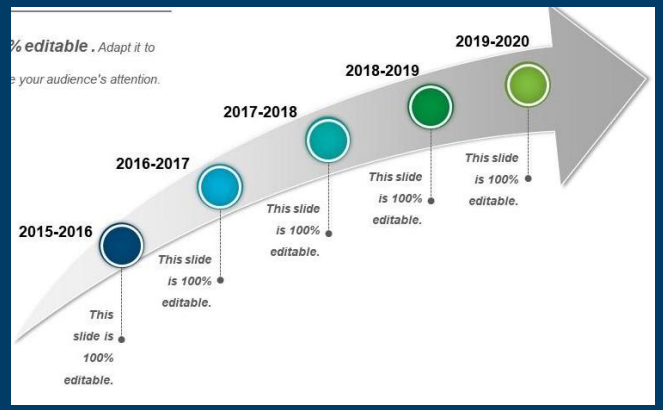
Phu Nguyen (phu.nguyen@sintef.no)
Seniorforsker, SINTEF Digital
https://www.linkedin.com/in/nguyenhongphu/


Teknologi for et bedre samfunn

# Outline

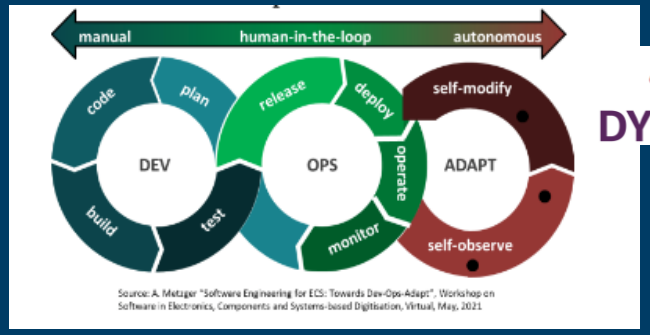
## 1-My (security) research

- About me
- My research roadmap




## 2-SOAR4BC

- (AI-driven) Security Orchestration, Automation and Response for Business Continuity of Critical Infrastructures




DYNABIC

## 3-Lessons learnt

- Explainable AI for Security

# Phu ("Foo") H. Nguyen

- Currently: **Senior Research Scientist** at SINTEF, Norway

- Previously :
  - **Research Scientist** at SINTEF, Norway.
  - **Postdoctoral Fellow** at Simula, Norway.
  - **PhD in Computer Science (Software Engineering)** at University of Luxembourg, Luxembourg.
  - **Master in Computer Science and Engineering** at Eindhoven University of Technology, The Netherlands.
  - **Bachelor in Computer Science** at Hanoi University of Science and Technology, Vietnam.
  - **Software Developer** at FPT, IBM Vietnam

Two weeks fresh in Norway, thought of hiking nearby Oslo with a group but ended up at Trolltunga!

# «How much AI» in my research projects so far???

# My research projects

## PhD project

- Model-Driven Security with Modularity and Reusability for Secure Systems Development

## Postdoc Projects

- Model-Based (Security) Testing for Cyber-Physical Systems
- Testing Cyber-Physical Systems under Uncertainty (**U-Test**)

## SINTEF

- **Cirrus** - Custom Code for Multi-tenant Cloud Computing
- Pilot-T **ASAM** project
- **ENACT**: Development, Operation, and Quality Assurance of Trustworthy Smart IoT Systems

## SINTEF

- **DAT4.Zero**
- **InterQ**: Interlinked process, product & data quality framework for zero defect manufacturing
- **REED**

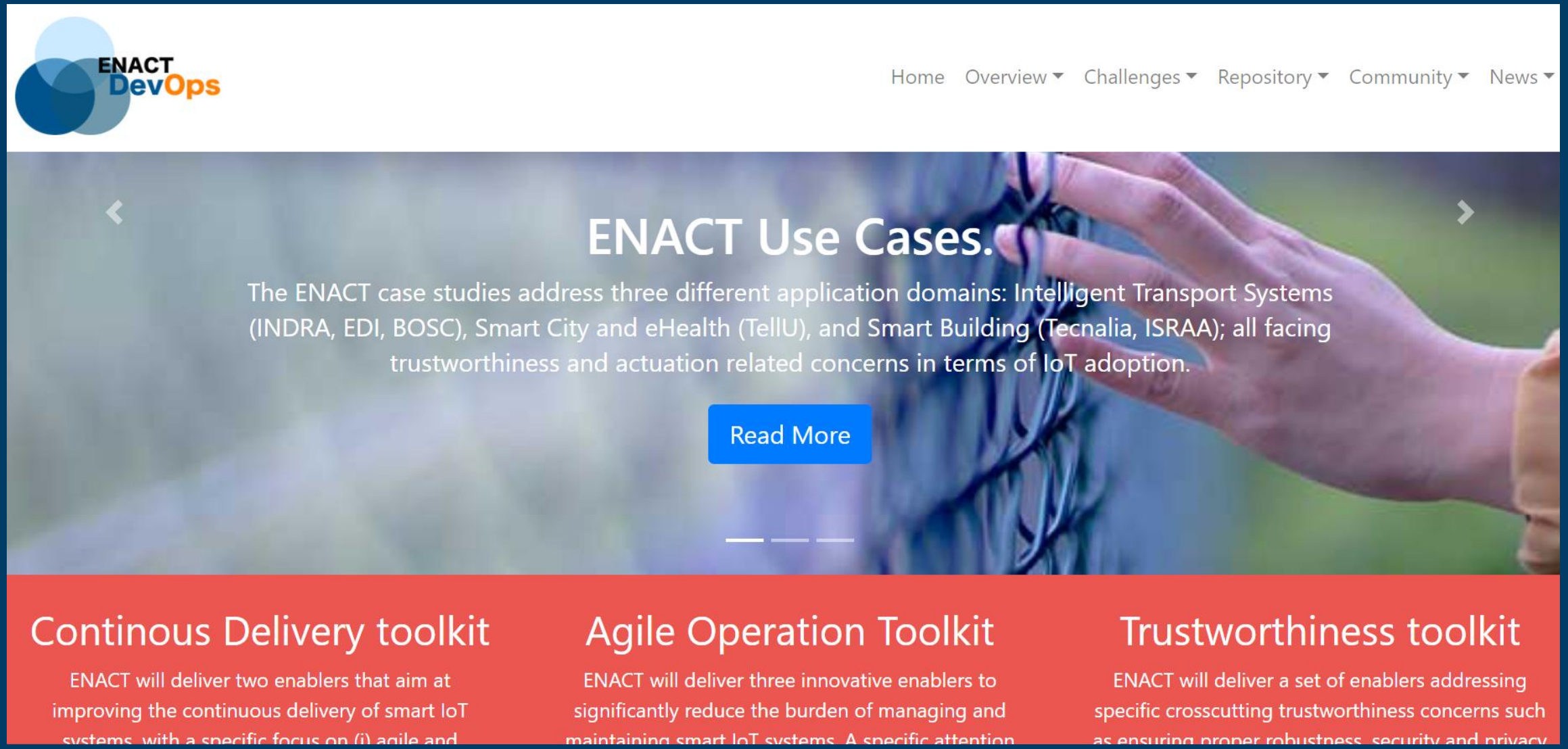
## SINTEF

- **DYNABIC**: Dynamic business continuity of critical infrastructures on top of adaptive multi-level cybersecurity
- **TechDebtOps**: Data-driven continuous management of technical debts for sustainable software development

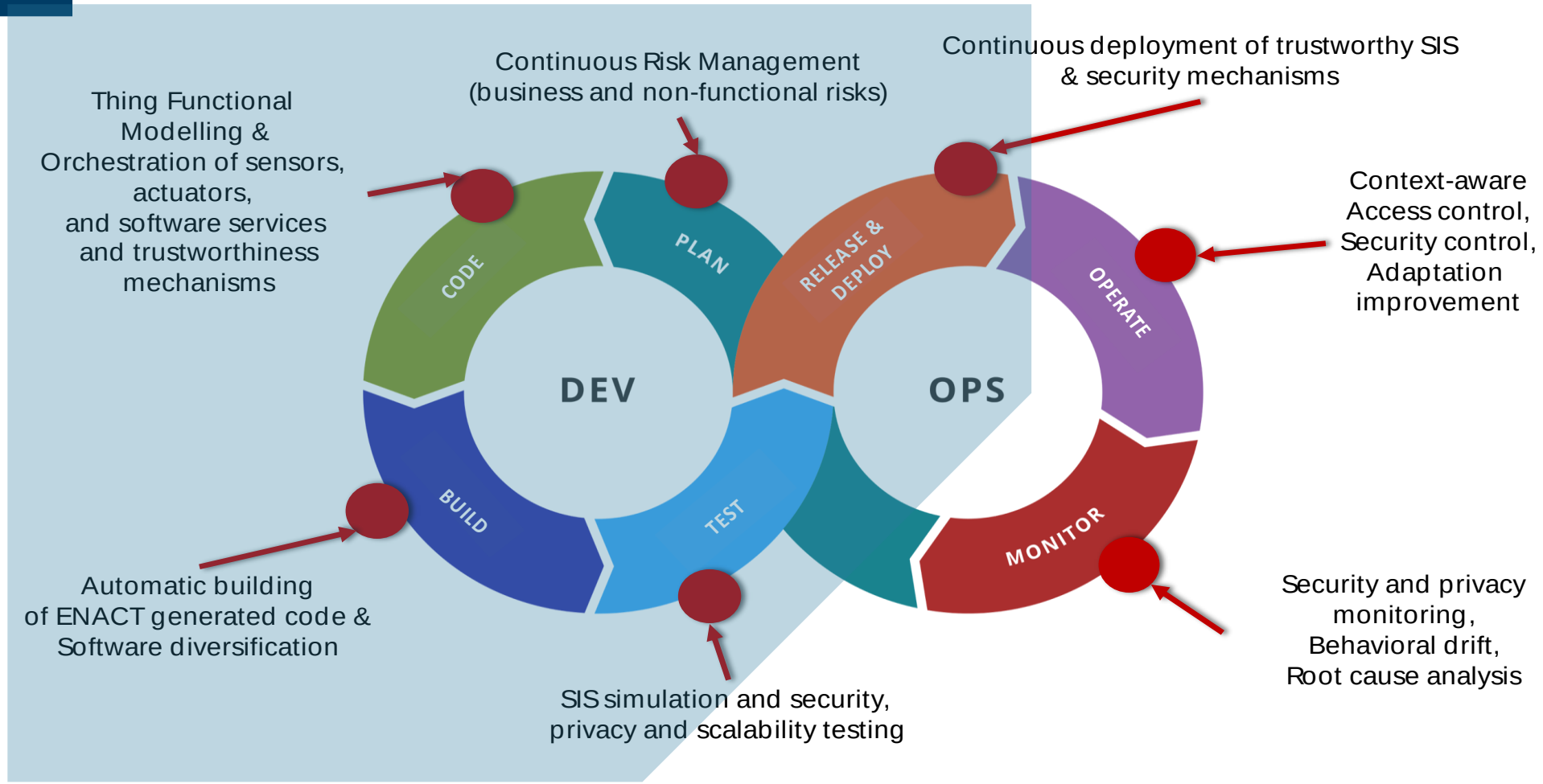
**AI, AI, & AI**

SINTEF

# Enabling DevOps for IoT



# ENACT project (EU project, 2018-2020): Development, Operation, and Quality Assurance of Trustworthy Smart IoT Systems (SIS)

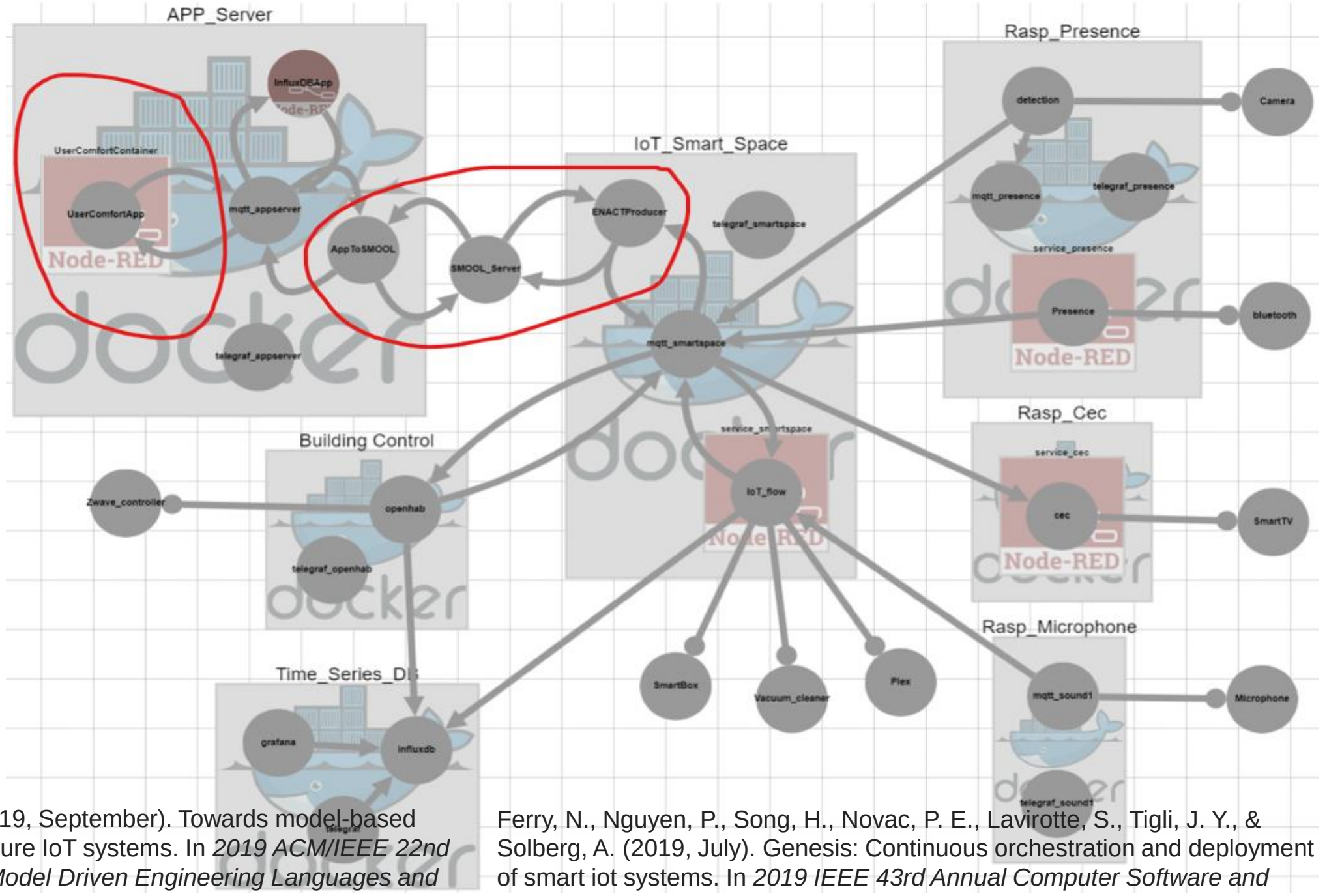



**Our SINTEF project team have developed tools to support the continuous agile modelling, development, testing and deployment of trustworthy SIS.**

# Lowcode

# GeneSIS tool for

# Continuous Orchestration and Deployment of Smart IoT Systems

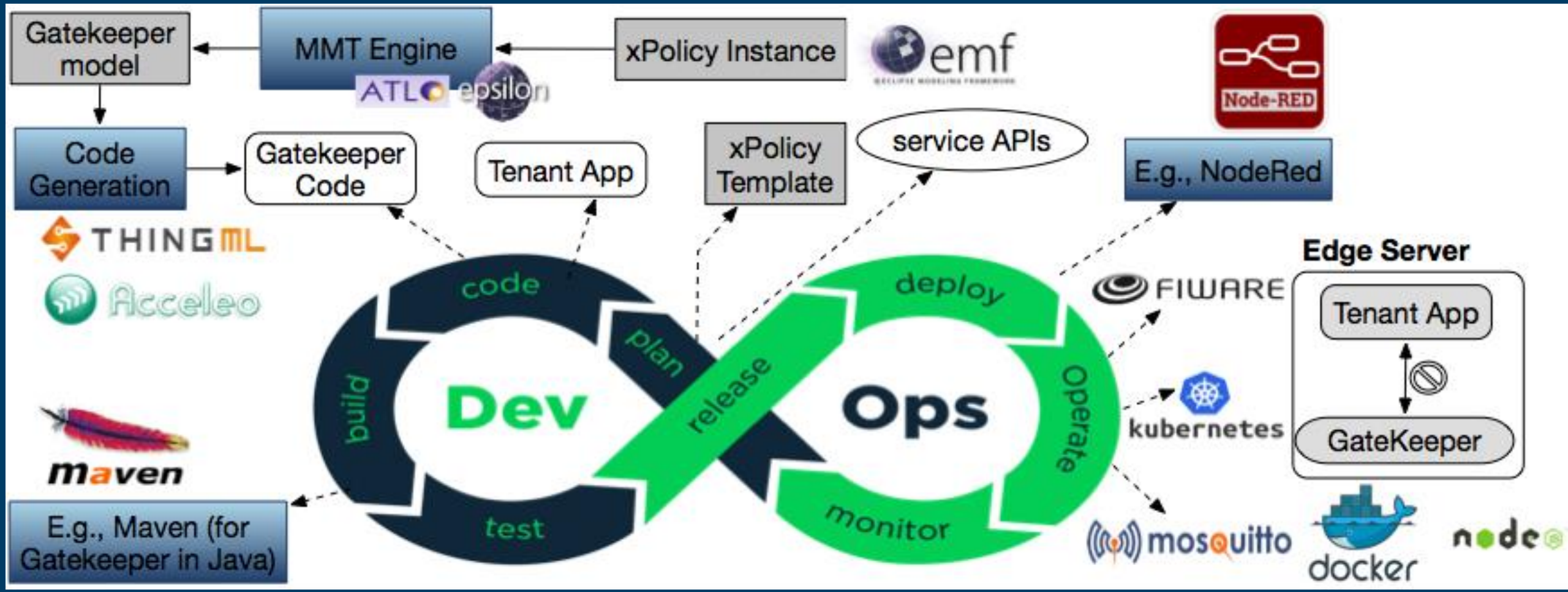




Ferry, N., & Nguyen, P. H. (2019, September). Towards model-based continuous deployment of secure IoT systems. In *2019 ACM/IEEE 22nd International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)* (pp. 613-618). IEEE.

Ferry, N., Nguyen, P., Song, H., Novac, P. E., Lavirotte, S., Tigli, J. Y., & Solberg, A. (2019, July). Genesis: Continuous orchestration and deployment of smart iot systems. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)* (Vol. 1, pp. 870-875). IEEE.

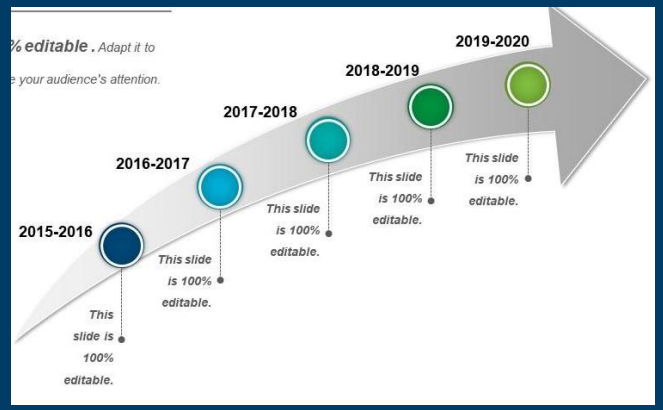
# MDSIoT in DevOps



# Outline

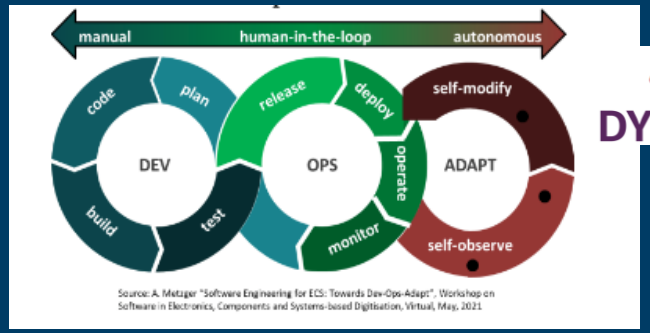
## 1-My (security) research

- About me
- My research roadmap




## 2-SOAR4BC

- (AI-driven) Security Orchestration, Automation and Response for Business Continuity of Critical Infrastructures



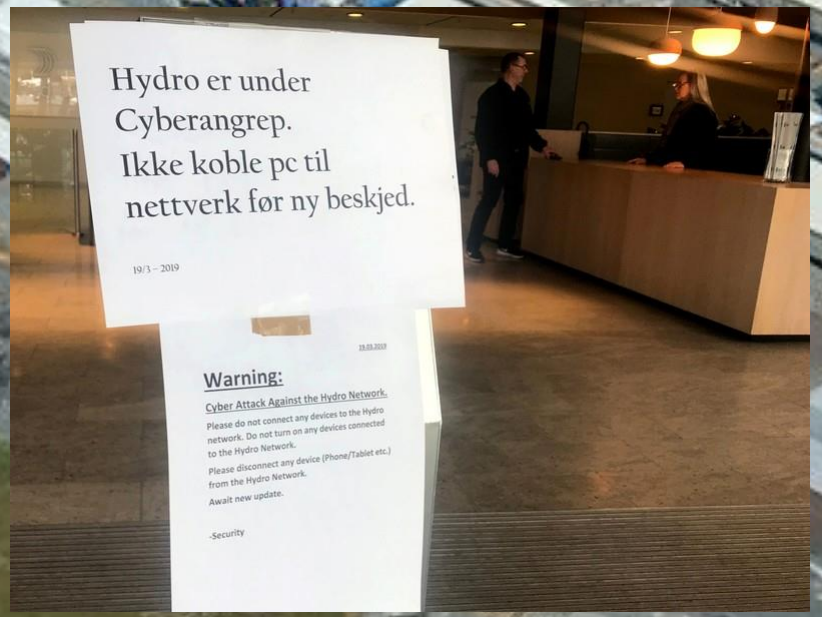

DYNABIC

https://www.hydro.com/en/global/media/on-the-agenda/cyber-attack/

# Hackers hit Norsk Hydro with ransomware

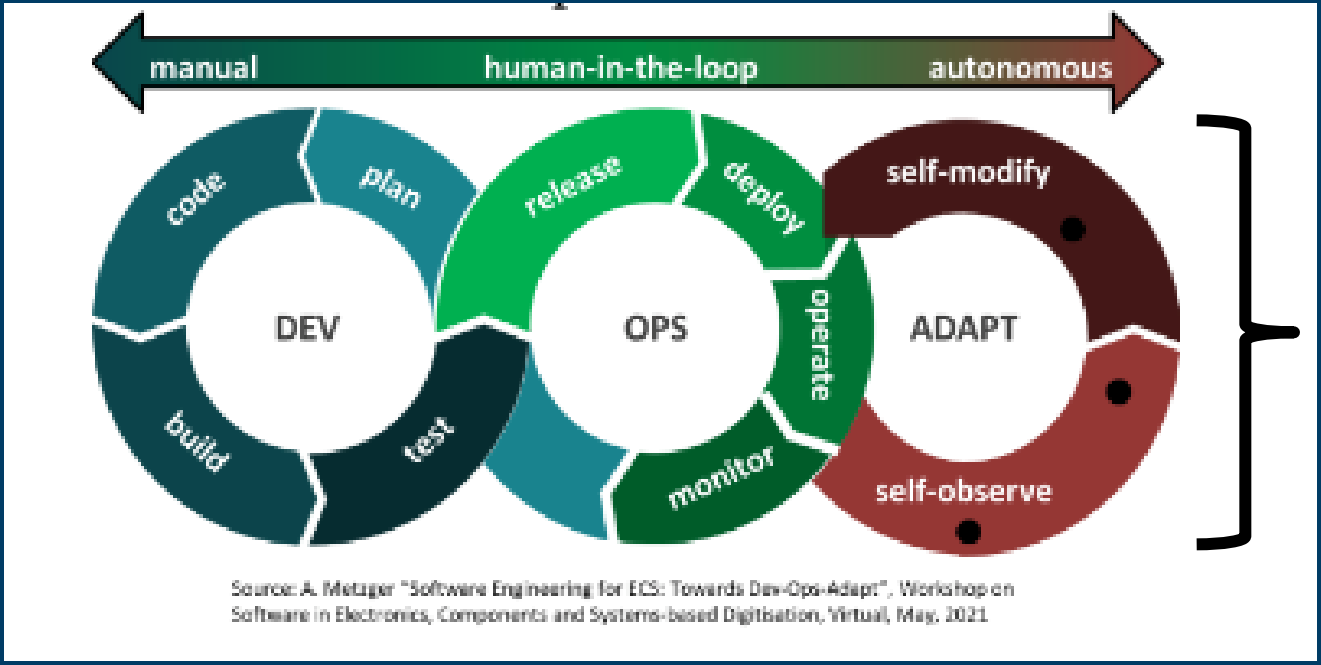
*December 2019*

**The financial impact would eventually approach $71 million.**



# Our DYNABIC project (EU project, 2022-2025)

- Business continuity risk management in critical infrastructures, based on **SecDevOpsAdapt cycle**




Source: A. Metzger "Software Engineering for ECS: Towards DevOps-Adapt", Workshop on Software in Electronics, Components and Systems-based Digitisation, Virtual, May. 2021

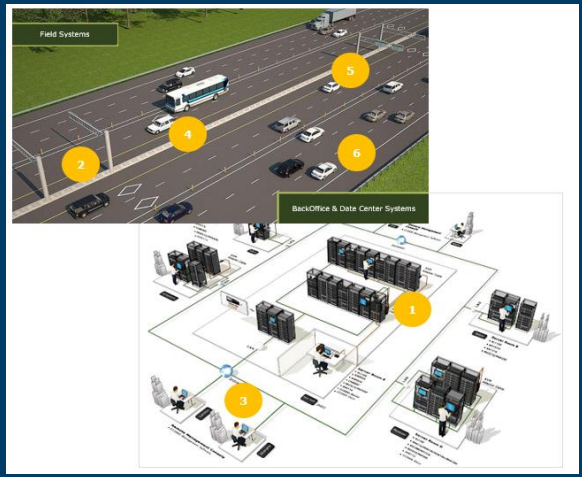
to enable the
**dynamic adaptation of the response to incidents and disruptions**.

https://dynabic.eu/

Erkuden Rios, Eider Iturbe, Angel Rego, Nicolas Ferry, Jean-Yves Tigli, Stéphane Lavirotte, Gerald Rocher, Phu Nguyen, Hui Song, Rustem Dautov, Wissam Mallouli, and Ana Rosa Cavalli. 2023. The DYNABIC approach to resilience of critical infrastructures. In Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23). Association for Computing Machinery, New York, NY, USA, Article 136, 1–8. https://doi.org/10.1145/3600160.3605055

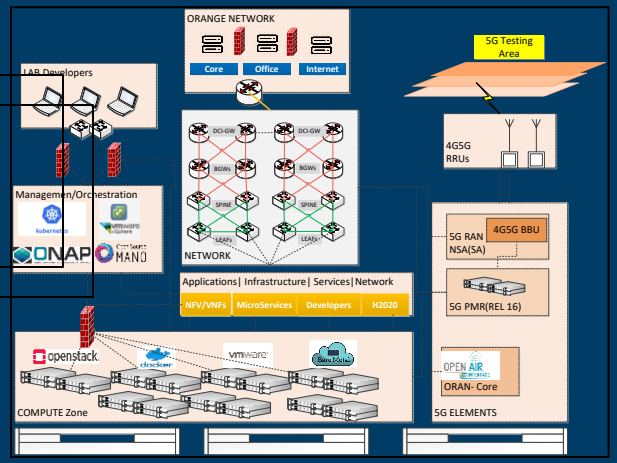
# How can we protect the critical systems for our society?

- Smart Preparedness, Prevention and Response to Business Disruption risks in critical infrastructures (CI) and supply chains

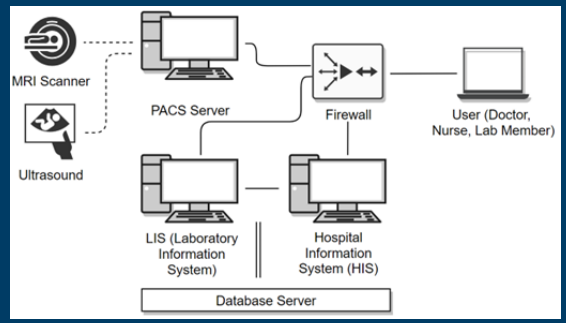



Transport services

EV charging station




5G Telco

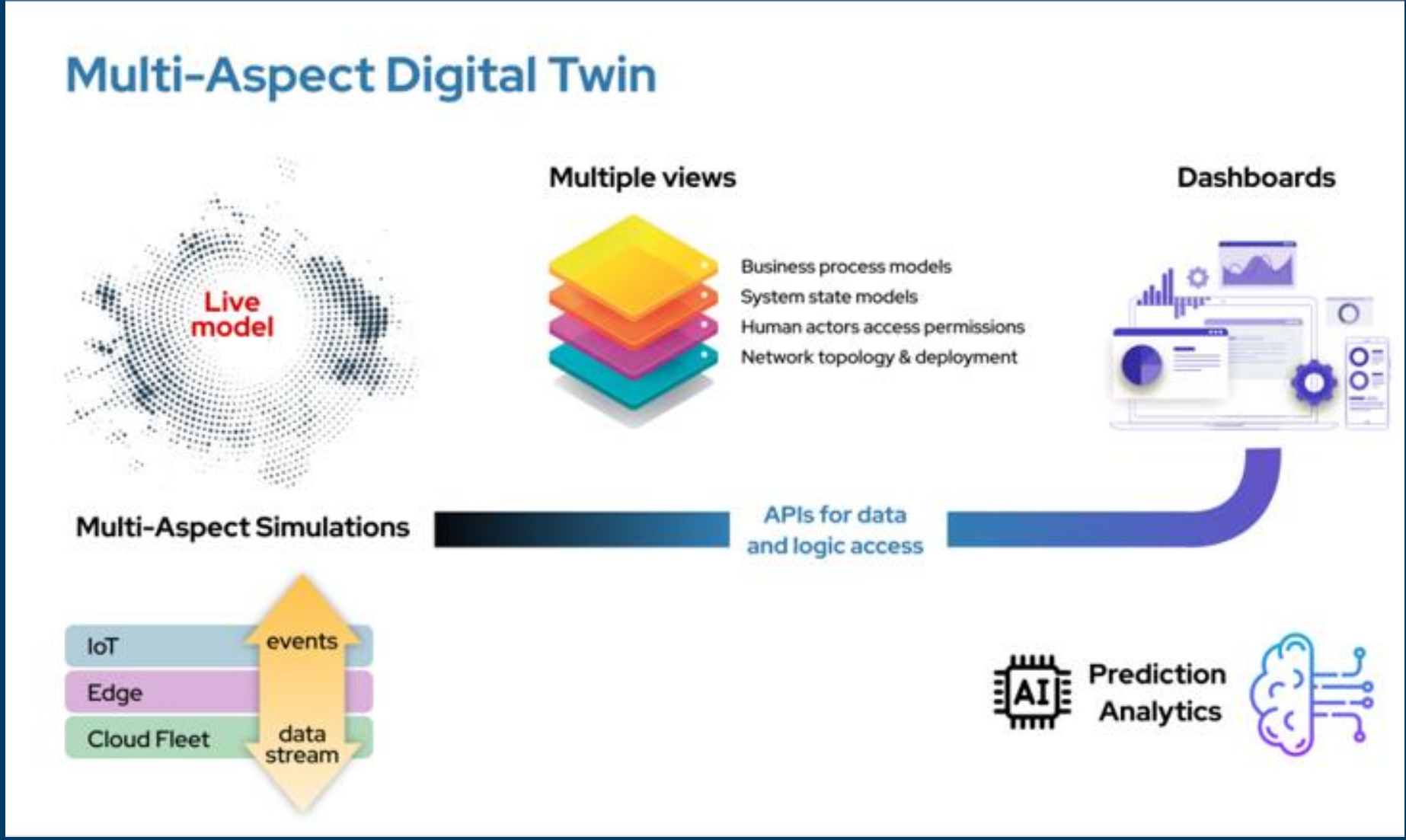



Health

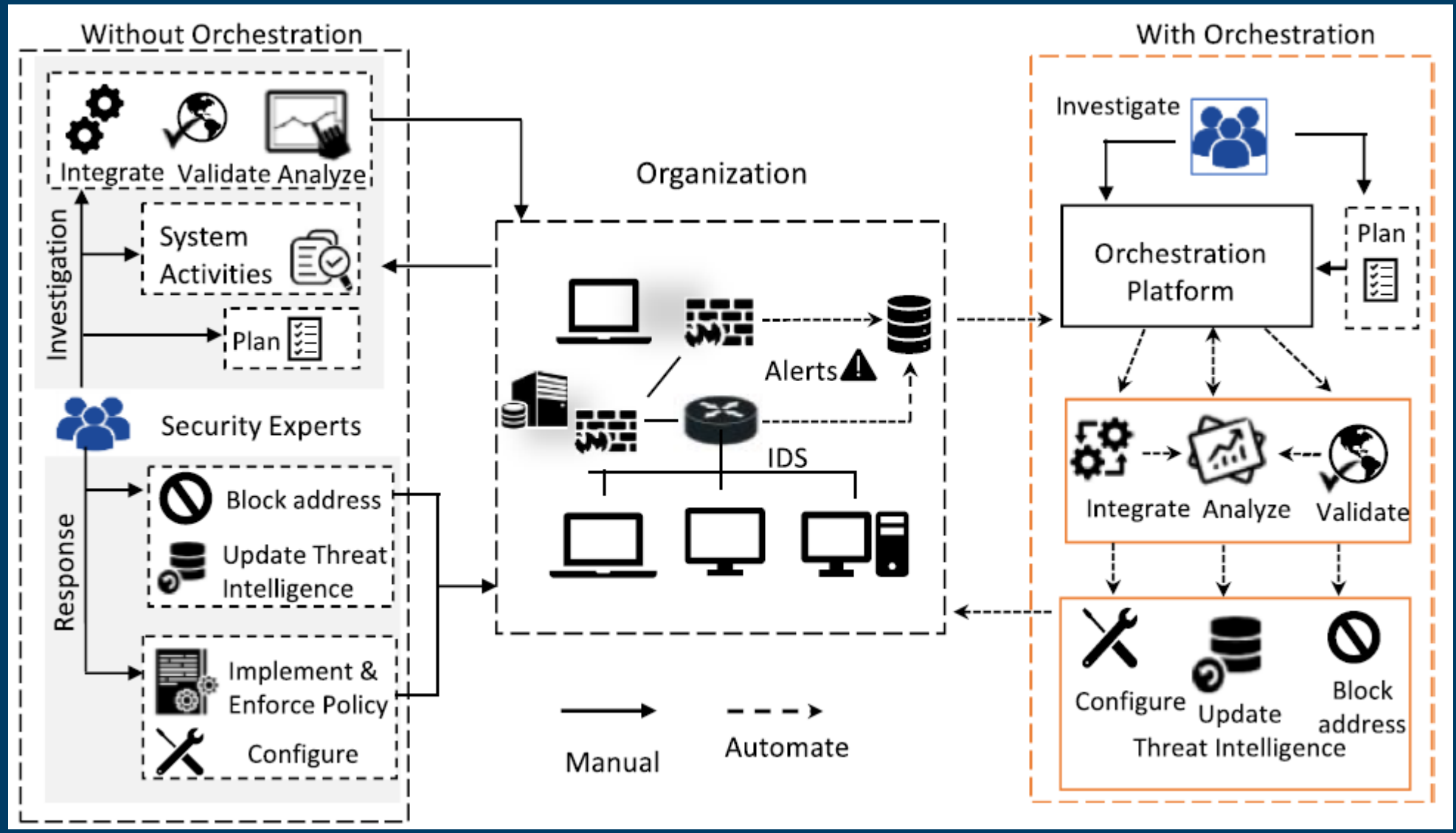
**Cascading effects in Interconnected CI**

# DYNABIC in a Nutshell

- A **Multi-Aspect Digital Twin** (MADT) of interconnected CIs.
- **Resilience solutions** on top of MADT to **analyse and predict potential business disruptions and their propagation.**



# The Transition to SOAR




Manual and passive processes of handling security incidents (On the left-hand side).
To the right: **More proactive and automated with (AI-driven) SOAR. Pros go with cons!**

Islam, C., Babar, M. A., & Nepal, S. (2019). A multi-vocal review of security orchestration. *ACM Computing Surveys (CSUR)*, *52*(2), 1-45.

**How to systematically unify different security mechanisms/tools and the assets in the CI system into SOAR?**

**How to to enable continuous security enhancement across various CI system levels?**

**How to develop advanced continual learning approaches to have more automation, but still highly explainable for human-in-the-loop?**

**Specialised SOAR solutions** from big tech companies such as Microsoft, IBM, Cisco, Fortinet, FireEye **are vendor lock-in**, **not flexible** for integrating with different security tools, especially with the physical assets across various CI system levels (e.g., via a DT).

This lack of flexible integration leads to **less systematic, less explainable SOAR solutions**, especially for cascading effects of CIs

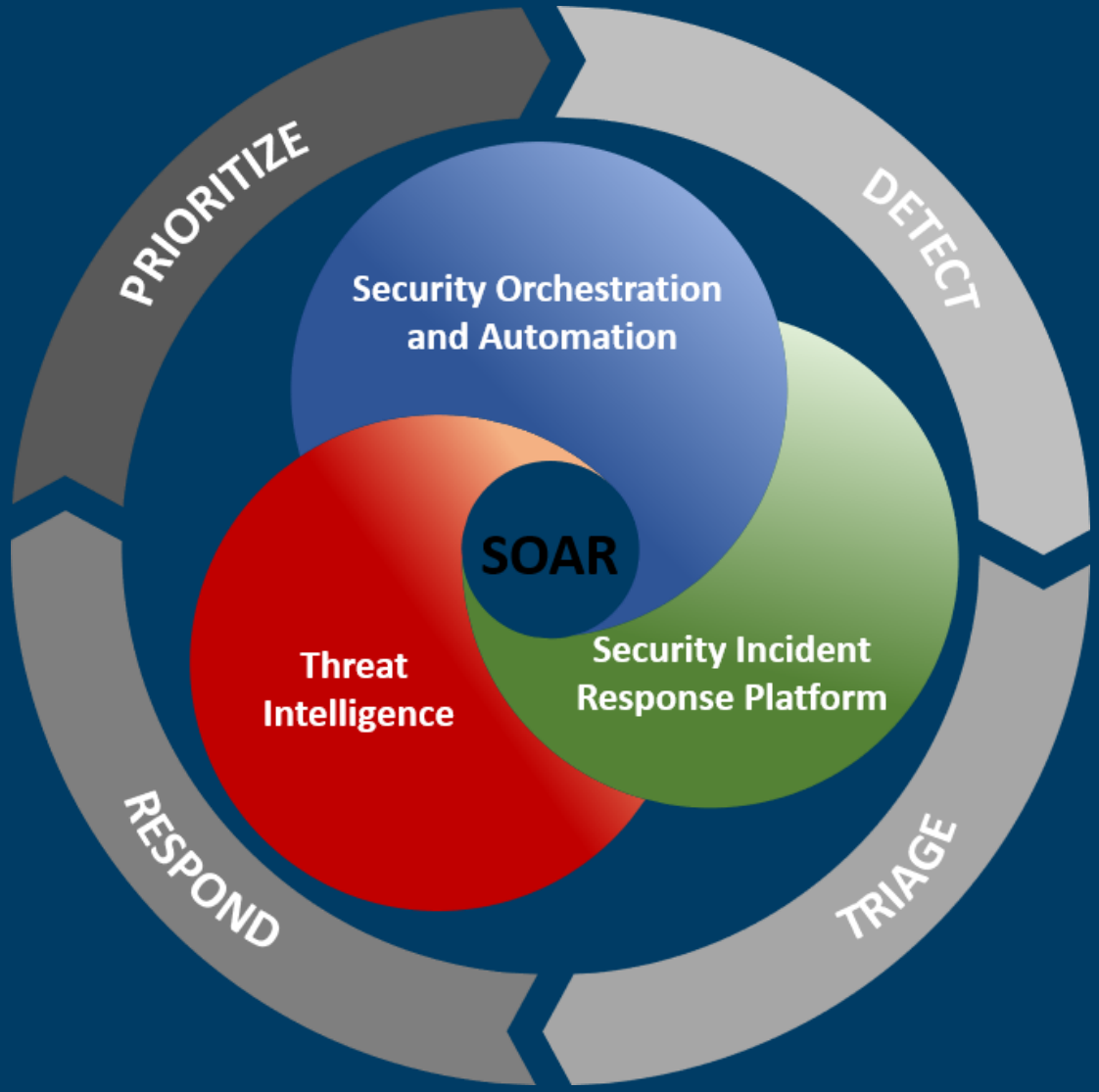
SINTEF

# The Challenges

**Critical infrastructure systems** currently **lack SOAR solutions embracing SecDevOps practices** with holistic **security context** as well as **system context** makes SOAR less systematic, especially for cascading effects.

**Lack of advanced continual learning** approaches to auto-generate SOAR playbooks dynamically in response to real-time threats.

**Lack of explainability** and advanced interfaces for **human-in-the-loop** support **during the automation and response** of SOAR solutions.




https://medium.com/@ddasmohapatra/security-orchestration-automation-and-response-soar-a7d929ad73ff

# SOAR4BC

SOAR4BC differentiates from the existing solutions by

1) empowering Reinforcement Learning (RL)-based **adaptation intelligence**,
2) leveraging the **system context from the digital twin** to provide a **holistic** security orchestration with **explainability**, and
3) **embracing SecDevOps practices** for security engineering of multi-layered systems (IoT/CPS) and considering the computing continuum perspective.

P. Nguyen *et al*., "Towards Smarter Security Orchestration and Automatic Response for CPS and IoT," *2023 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Naples, Italy, 2023, pp. 298-302, doi: 10.1109/CloudCom59040.2023.00055.

**Reinforcement Learning (RL)-based adaptation intelligence** orchestrates a combination of automatic and human responses

**Security knowledge graph** as the extension of the platform's knowledge graph to connect different security tools, system assets with the complex data source and models

Integrated and customizable **dashboard** for multi-aspect visualization of the SOAR operations and analysis results

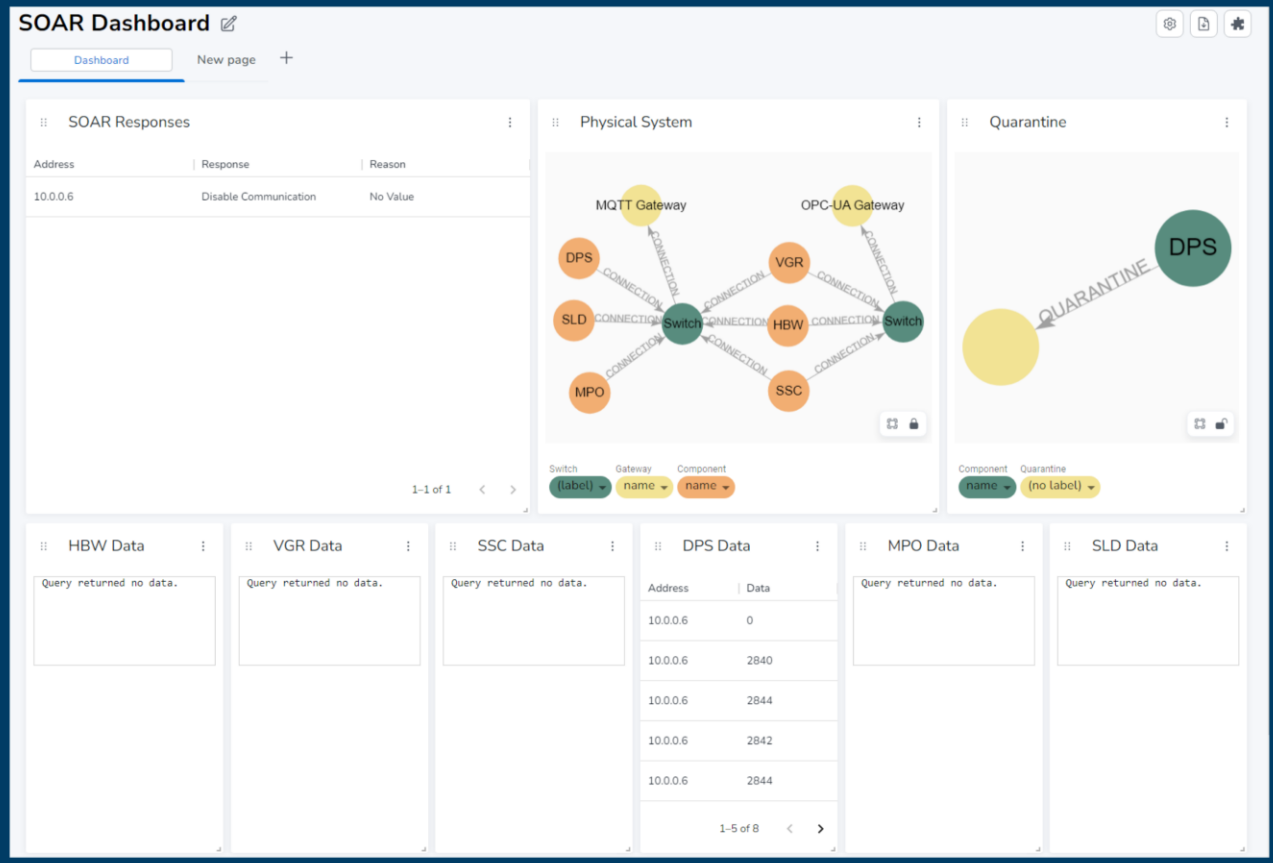
**SecDevOps** practices are enabled as part of SOAR for the **continuous enhancement of security** solutions aligned with the evolution of CIs

**Natural language interface** for explaining, supporting SOAR decision making with human-in-the-loop

# Technology

## Key technical features:

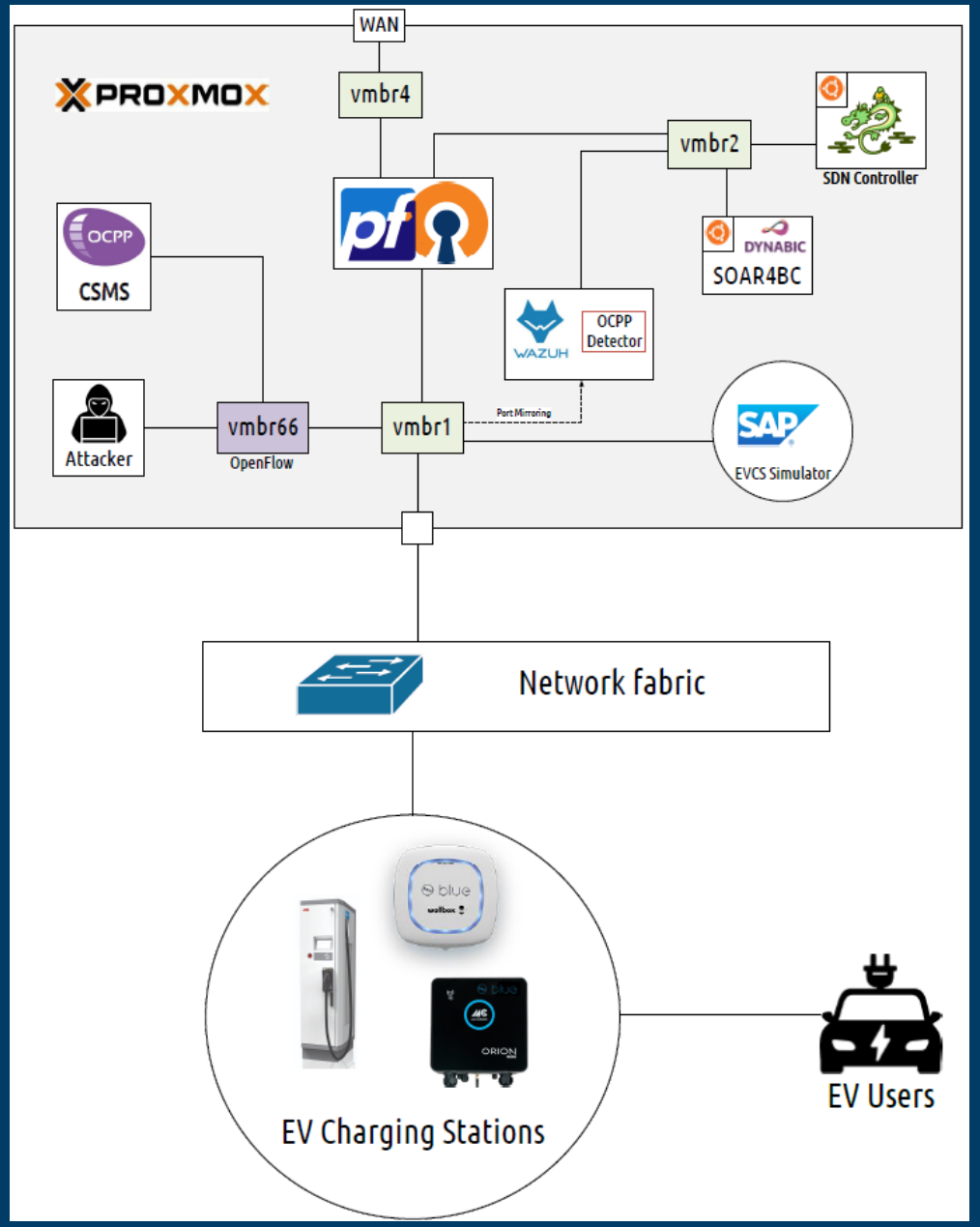
- Reinforcement Learning (RL)-based adaptation intelligence
- Security-oriented Digital Twin-based
- Knowledge graphs, object and time-series databases
- Container infrastructure
- Customizable dashboard
- Continuous IoT orchestration & deployment



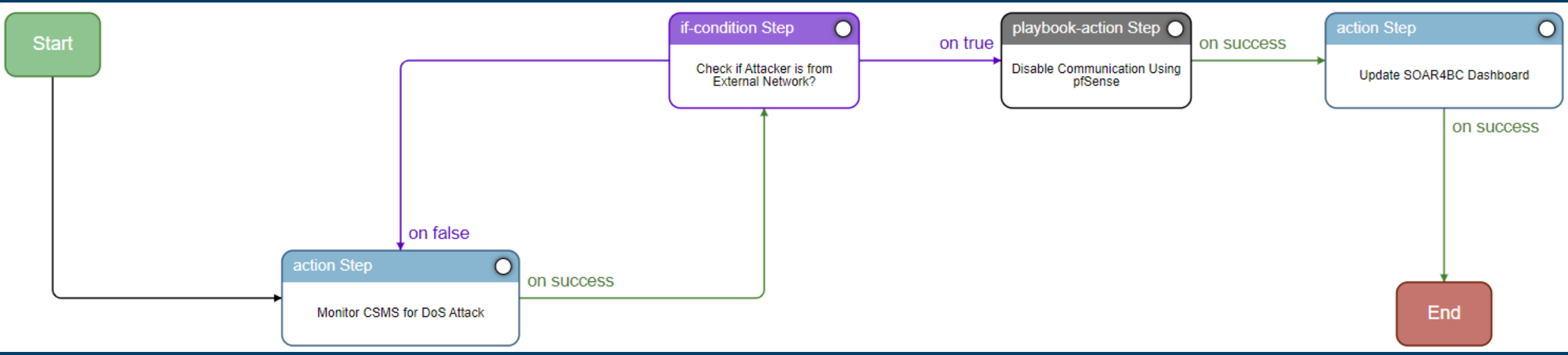
# The Public Power Corporation (PPC)'s EV Charging stations

- SOAR4BC for the Business Continuity against Security Threats to EV charging stations.
- Experimental scenarios:
  - False Data Injection attack
  - Denial-of-service (DOS) attack
- Results: New insights into the convergence of digital twin technology and cybersecurity with (explainable) AI-driven SOAR

*Testbed provided by Christos Dalamagkas (PPC)*



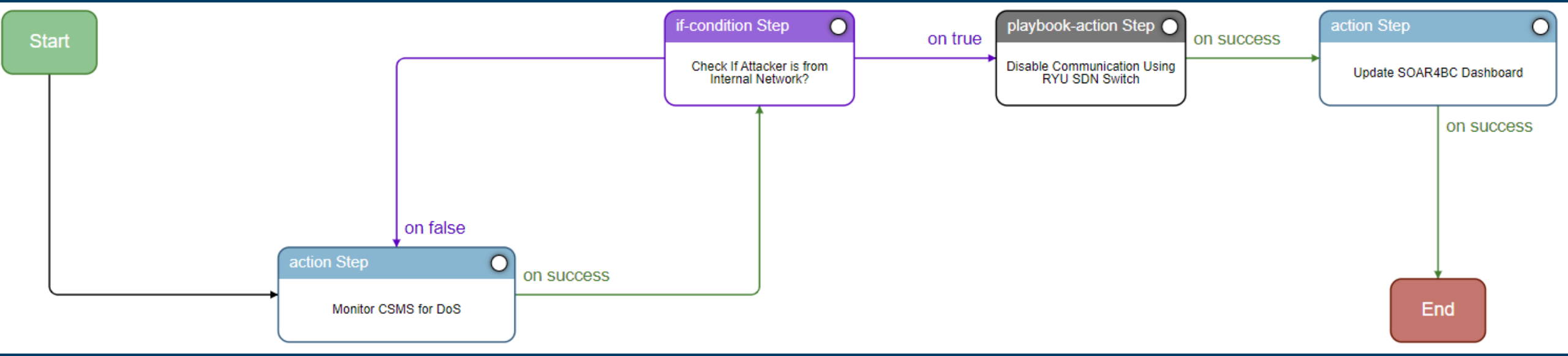
# CACAO v2 Playbook for DoS attack 1



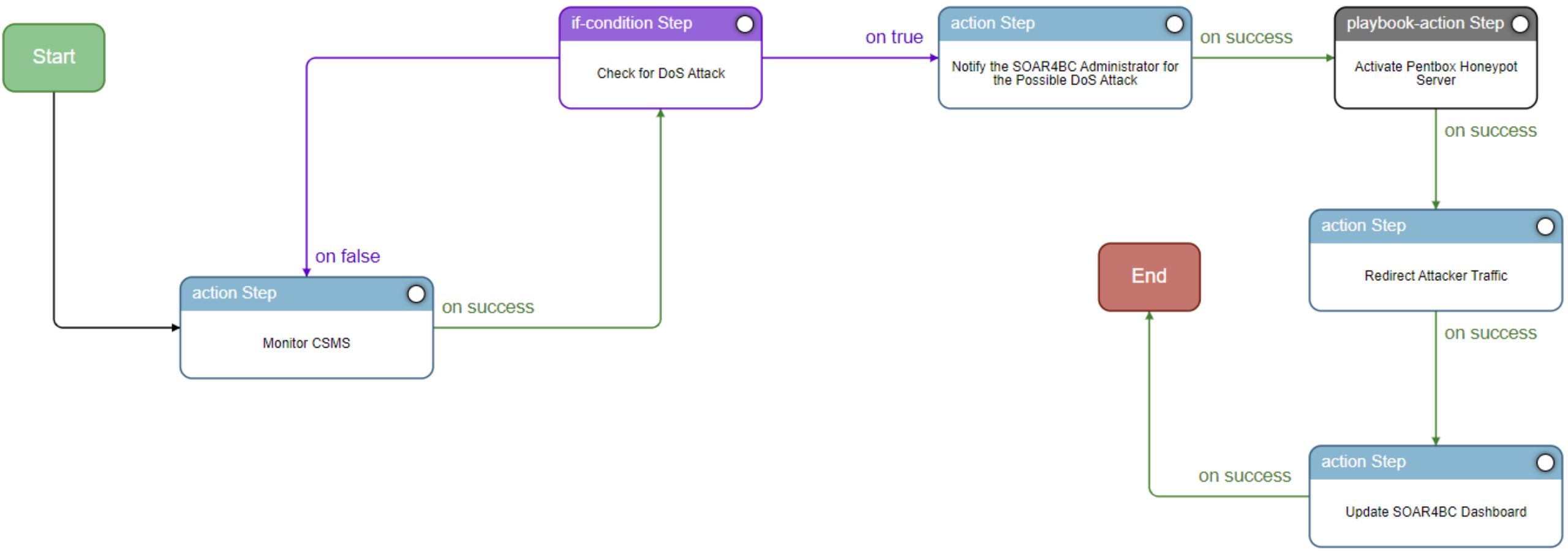

Playbook is “a standard set of operational procedures (playbook) to be used in planning and conducting a cybersecurity vulnerability and incident response activity.”

https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/

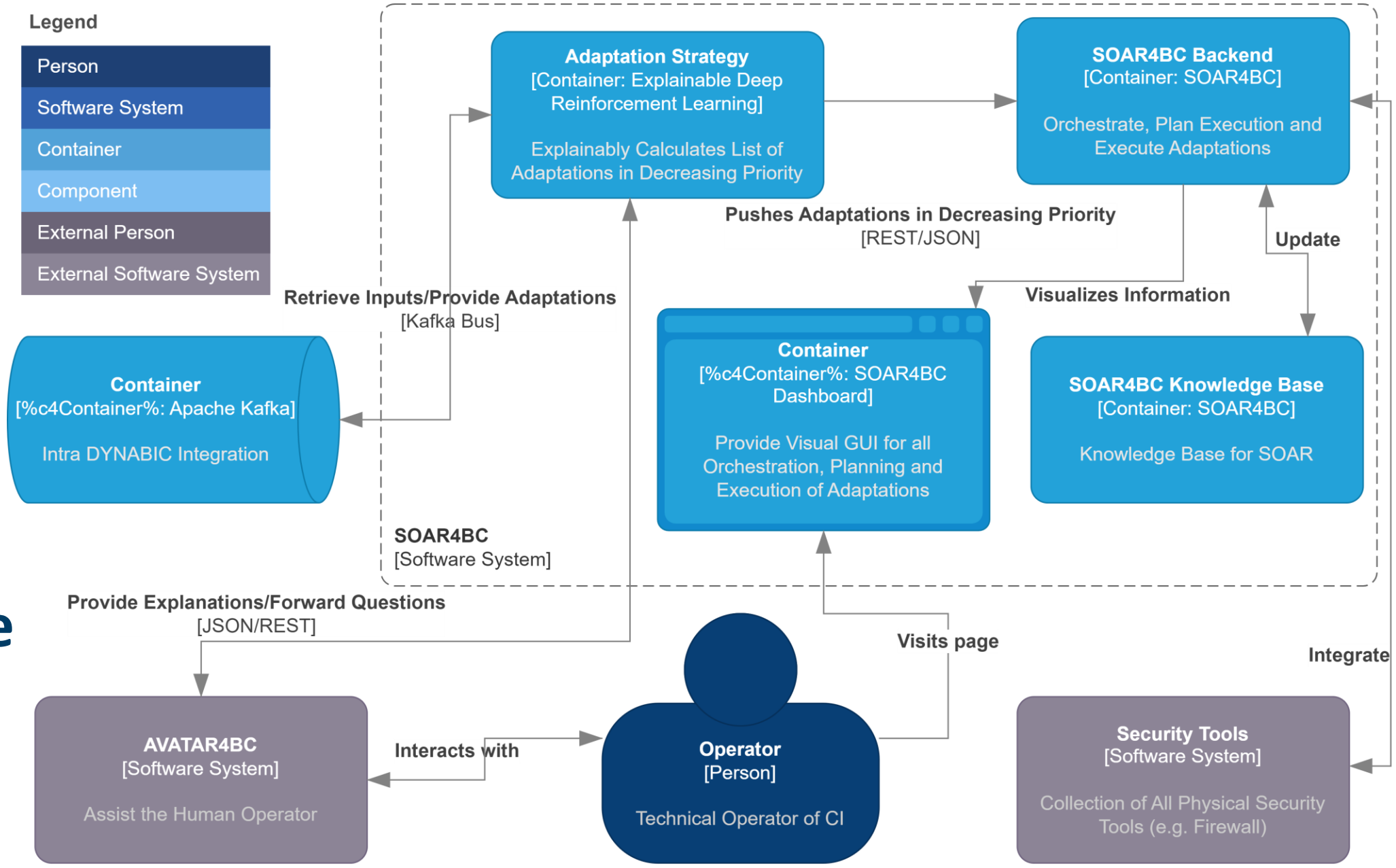
# CACAO v2 Playbook for DoS attack 2



# CACAO v2 Playbook for DoS attack 3



# SOAR4BC architecture



# Reinforcement Learning (RL)-based adaptation intelligence

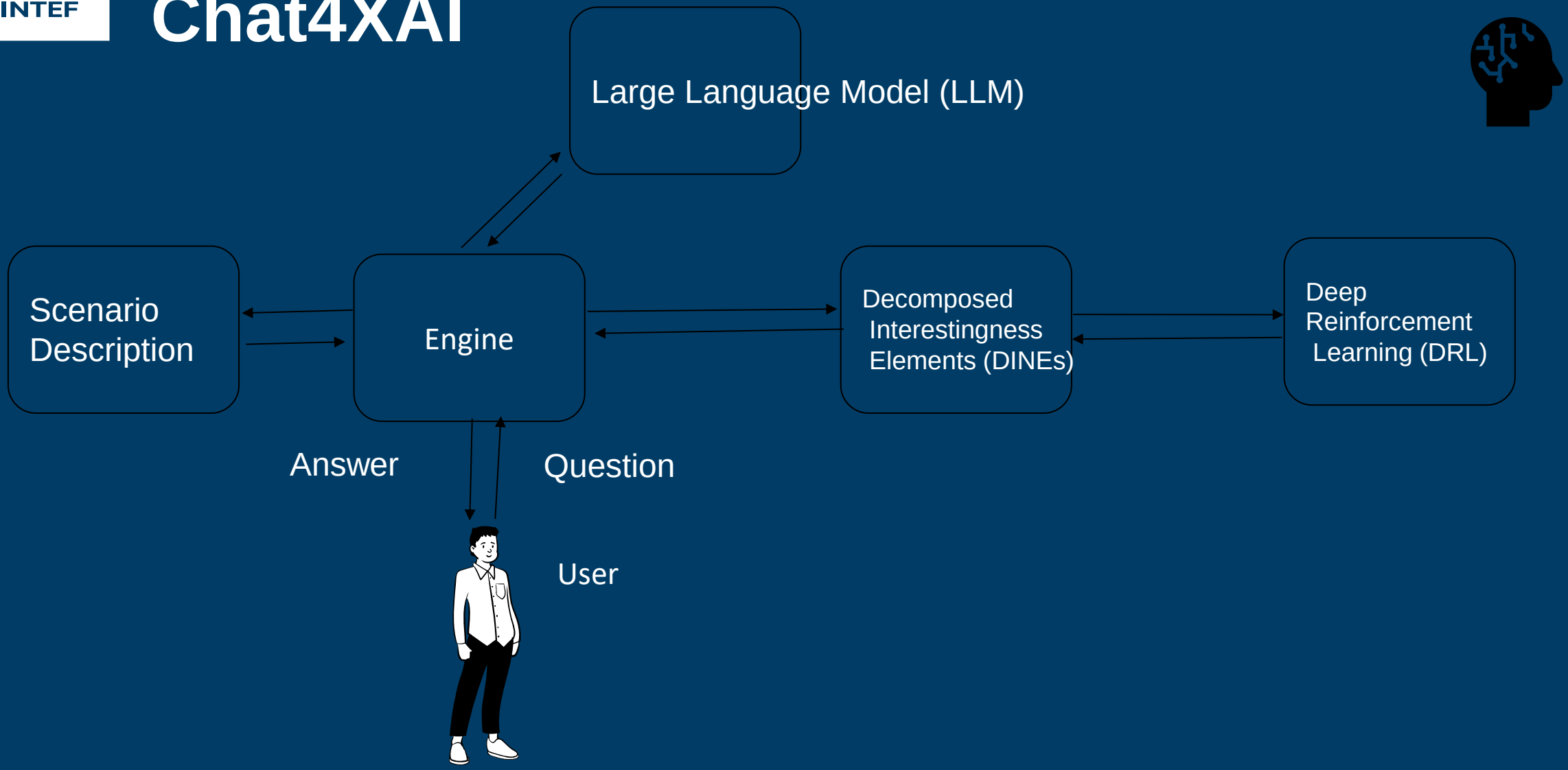
- **Problem**: Manually defining *when* and *how* CI must change their behaviour at runtime is challenging (i.e., design time uncertainty)
  - Incomplete knowledge about (1) all future environmental situations & (2) effects of a particular change
- **Solution**: Online Deep Reinforcement Learning
  - RL architecture based on Deep-Q-networks with experience replay
  - Reward function uses business goals: availability, profit, image
- **Result**: Runtime security orchestration based on live data

# Explainability of RL-based adaptation intelligence

- Human-in-the-loop design
  - Human can observe decision-making of RL-based adaptation intelligence
  - Human can intervene in critical situations
- **Realisation**: Chat4XAI, a natural language chatbot [1]
  - Input: Natural language questions
  - Processing: Uses Decomposed Interestingness Elements (DINEs) and Large Language Model (LLM) (i.e., GPT4, LLM is interchangeable)
  - Output: Natural language explanations
- **Example** (from the PPC use case):
  - Question: "*Why was the action selected on timestep 499?*"
  - Answer: "*The action <<do nothing>> was selected because* [...] *it would have a higher positive impact on the goal of Revenue compared to the action <<Spin up honeypot server>>* [...]*, which would maximize the goals of Availability and Image. The agent decided to prioritize Revenue in this instance."*

[1] Andreas Metzger, Jone Bartel and Jan Laufer. “An AI Chatbot for Explaining Deep Reinforcement Learning Decisions of Service-Oriented Systems”. International Conference on ServiceOriented Computing (ICSOC). Springer. 2023, pp. 323–338.

# Chat4XAI

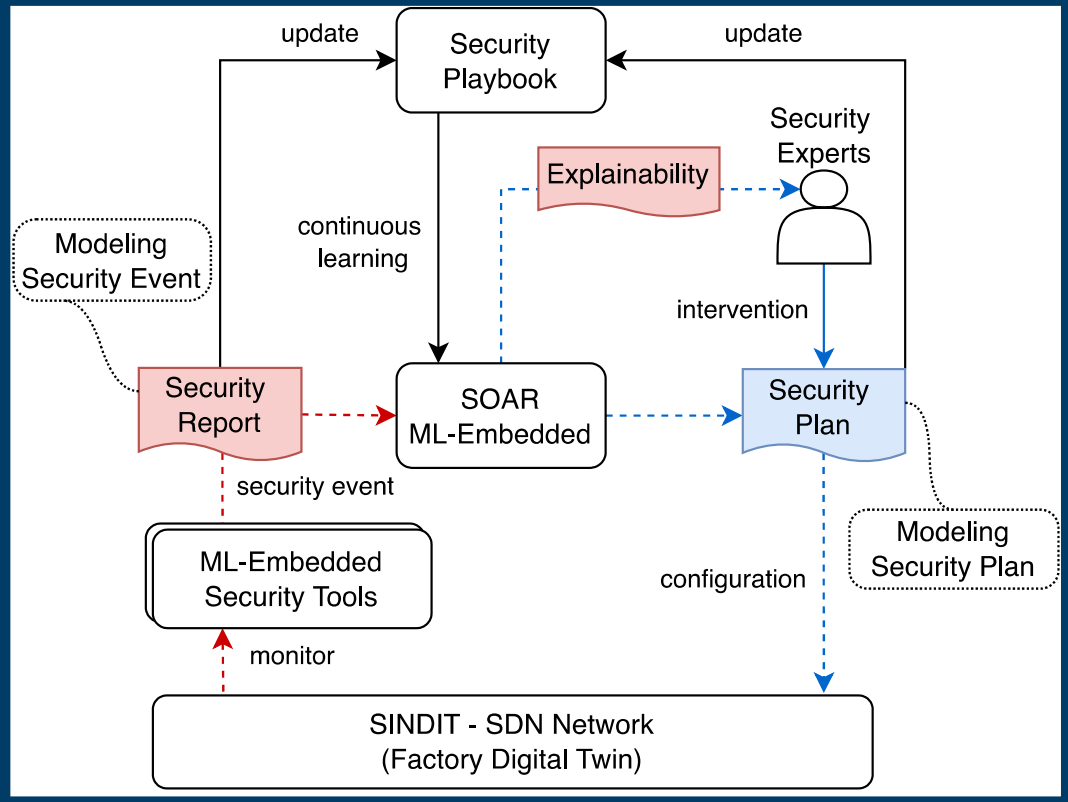



[1] Andreas Metzger, Jone Bartel and Jan Laufer. “An AI Chatbot for Explaining Deep Reinforcement Learning Decisions of Service-Oriented Systems”. International Conference on ServiceOriented Computing (ICSOC). Springer. 2023, pp. 323–338.

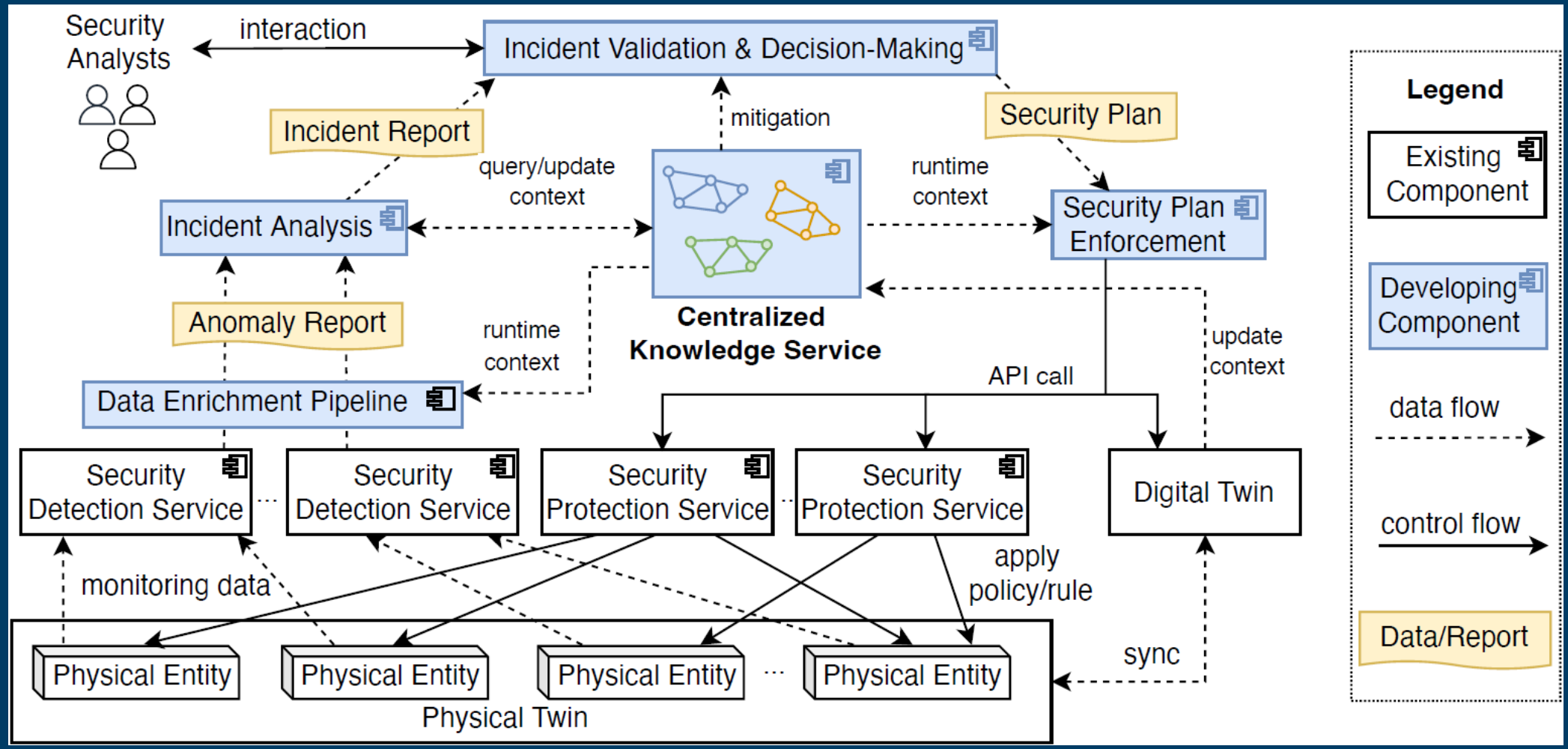
SINTEF

# Orchestration with explainability

- Building an Explainable Model for Orchestrating Security Services:
  - Supporting Security Tool Unification
  - Facilitating Security Playbook-based
  - Improving Explainablity in Orchestrating ML-based Security Services



# RXOMS - Runtime eXplainability for Orchestrating ML-base Security Framework




M. -T. Nguyen, A. N. Lam, P. Nguyen and H. -L. Truong, "Security Orchestration with Explainability for Digital Twins-Based Smart Systems," 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 1194-1203, doi: 10.1109/COMPSAC61105.2024.00159.

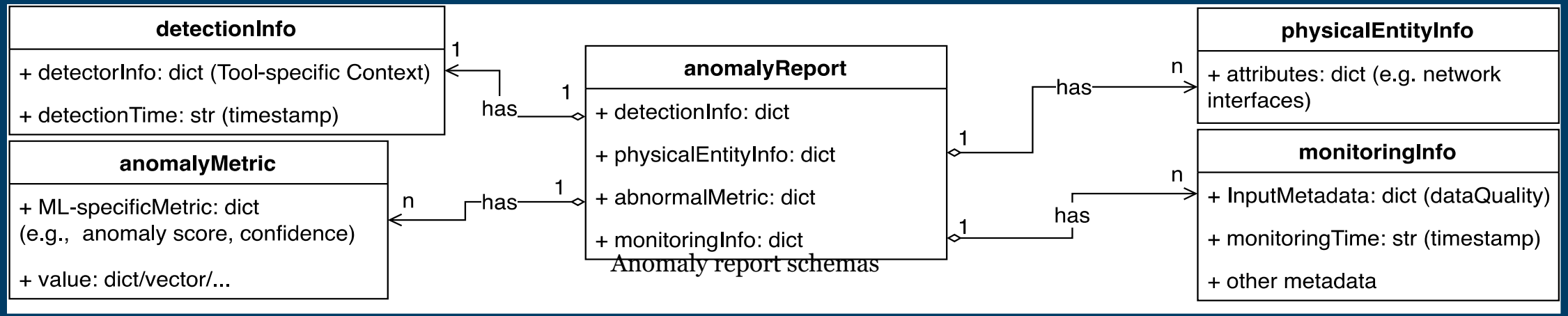
# Data Enrichment Pipeline

**Security Unification:**

- Capture runtime metrics and associated information at physical layer
- Enrich anomaly report from different security tools with comprehensive information and structure the information in a **unified schemas**

➢ Report anomalies within specific contexts

- **System context**: physicalEntityInfo, monitoringInfo
- **Tool-specific context**: detectionInfo
- **Security context**: anomalyMetric



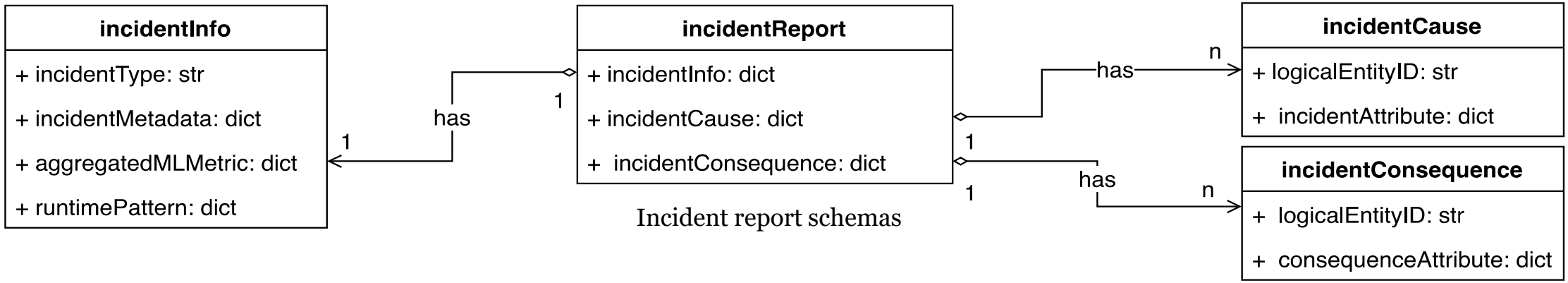
Anomaly report schemas

M. -T. Nguyen, A. N. Lam, P. Nguyen and H. -L. Truong, "Security Orchestration with Explainability for Digital Twins-Based Smart Systems," 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 1194-1203, doi: 10.1109/COMPSAC61105.2024.00159.

# Incident Analysis

## Runtime Explainability:

- Update runtime contexts from anomaly reports to the **Centralized Knowledge Service**
- Trace and analyze incidents based on runtime contexts
- Report incidents via **logical entities**
- ➢ Report incidents with **enhanced explainability**
  - IncidentInfo: incident metadata, attacking method, pattern, and metrics
  - IncidentCause: attack origin, logical entity IDs
  - IncidentConsequence: attack targets/affected entities, logical entity IDs




Incident report schemas

M. -T. Nguyen, A. N. Lam, P. Nguyen and H. -L. Truong, "Security Orchestration with Explainability for Digital Twins-Based Smart Systems," 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 1194-1203, doi: 10.1109/COMPSAC61105.2024.00159.

# Runtime explainability in identifying and mitigating security incident

Vectra AI Incident log


```
{
  "UTCTimeStart": "1701461667", "UTCTimeEnd": "1701461852",
  "dd_dst_ip": "10.0.0.4",
  "dd_proto": "TCP",
  "category": "DDoS",
  "certainty": 72,
  ...}
```


Incident log with attributes from physical layer

RXOMS Anomaly Report


```
"reportID": "276852de-744a",
    "detectionInfo":{
        "detectionTime":"1701462137",
        "configuration": {"MLModel":{}, "threshold":{},...}
        ...},
    "physicalEntityInfo":{
        "0192eb5c-e33b": { # entityID
            "attribute": {
                "in_port": 2,
                "eth_dst": "10.0.0.4"
            },...}
    },
    "anomalyMetric":{
        "detectionConfidence": 0.72,
        "anomalyScore": -0.95,
        "byteCount": {
            "value": [486565],
            "unit": "Mbyte"}
    ...},
    "monitoringInfo":{
        "monitoringTime": "1701461667"},
        "dataQuality": {"consistency": 100,...}
        "sampleRate": 1,
    ...
```


RXOMS Incident Report


```
"reportID": "276852de-744a",
    "incidentInfo":{
        "incidentID": "T1498.002",
        "incidentName":"Network_Denial_of_Service",
        "runtimePattern": "HTTP Flood",
        "aggregatedMLMetric": {"minConfidence": 0.6, ...},
        ...},
    "incidentCause":{
        "47cf8b42-2eb4": {   # entityID
            "attribute": {
                "name": "mpo",
                "type": "factoryAsset"
                ...}}},
    "incidentConsequence":{
        "96er8s42-w2e6": {
            "attribute": {
                "name": "mqttGateway",
                "status": "disrupted",
                ...}},
        "0192eb5c-e33b": {
            "attribute": {
                "name": "S1",
                "type": "switch",
                "port": [2,3,5],
                "status": "overflow"}}
    },...
```


Anomaly and Incident are reported in specific contexts with **enhanced explainability on logical layer**


M. -T. Nguyen, A. N. Lam, P. Nguyen and H. -L. Truong, "Security Orchestration with Explainability for Digital Twins-Based Smart Systems," 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, 2024, pp. 1194-1203, doi: 10.1109/COMPSAC61105.2024.00159.

# COMPETITION

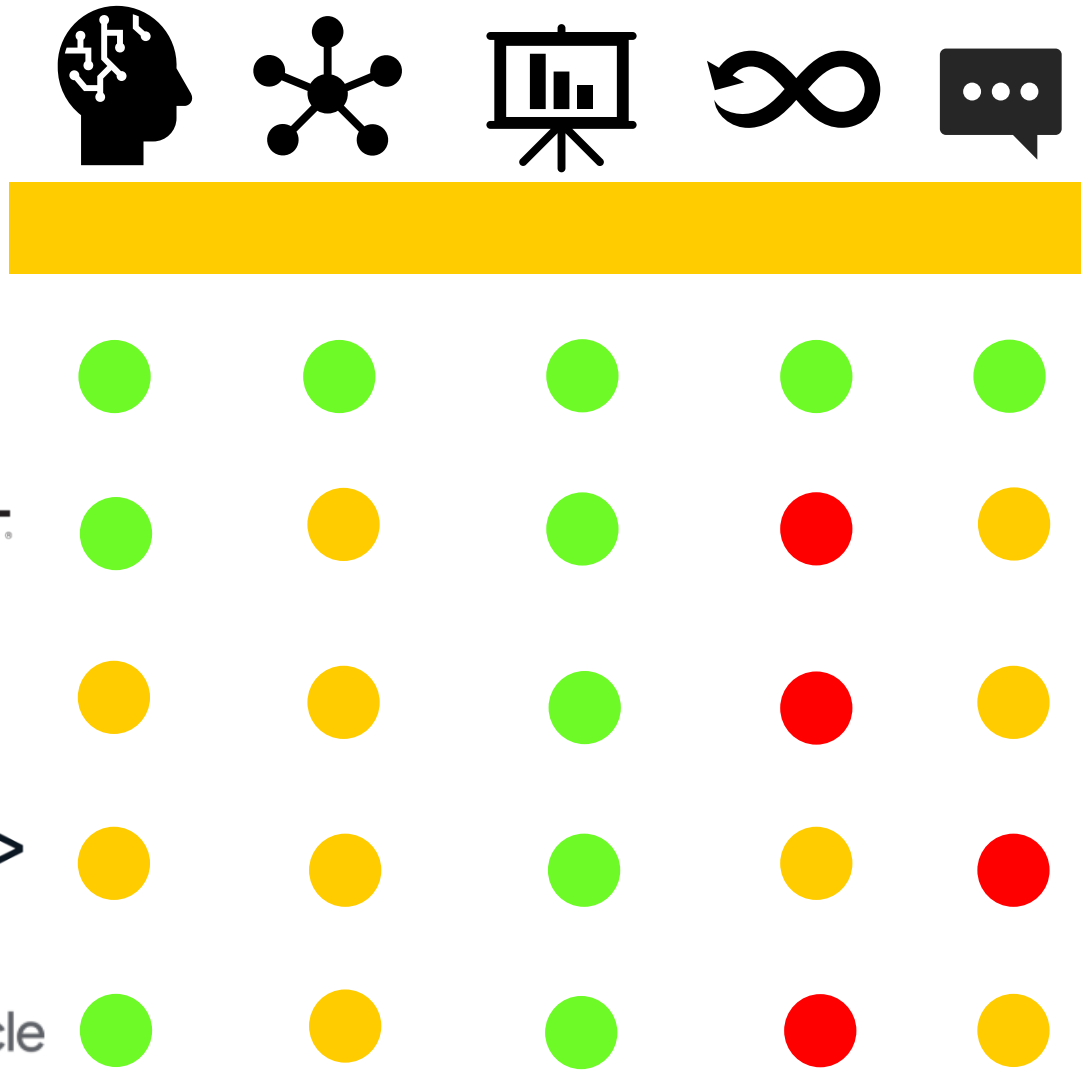
**Adaptation Intelligence**

**Holistic integration of security context and system context**

**Customizable dashboard**

**SecDevOps practices enabled**

**Automation with explainability for human-in-the-loop**



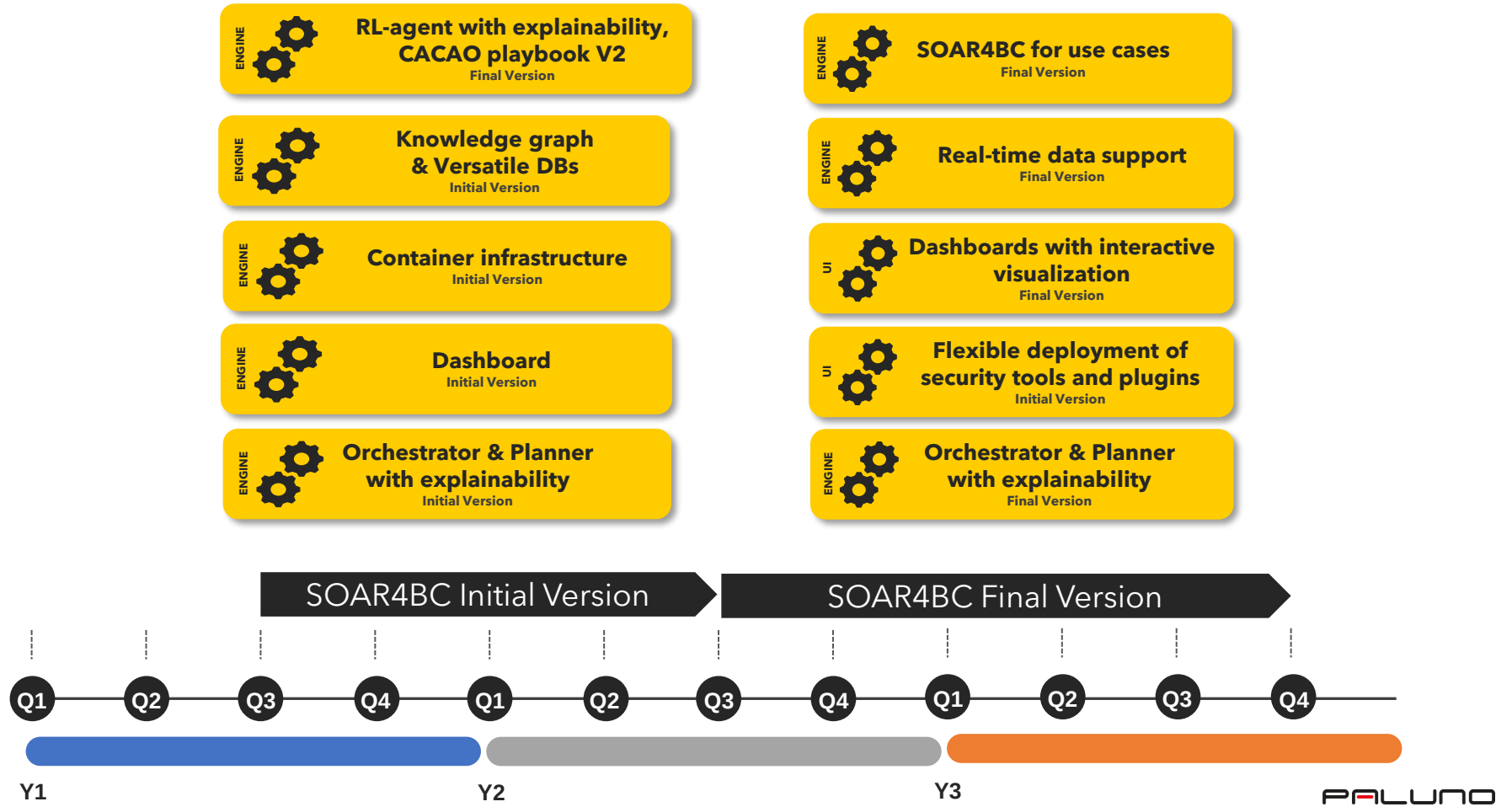

https://expertinsights.com/insights/the-top-soar-solutions/

SINTEF

PALUNO
The Ruhr Institute for Software Technology

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

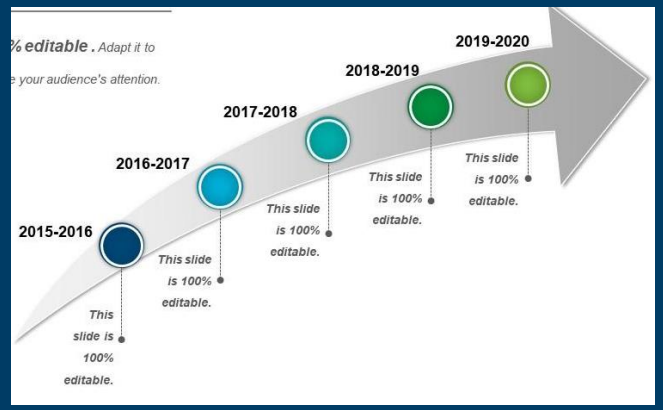
# Current status and technical roadmap



# Outline

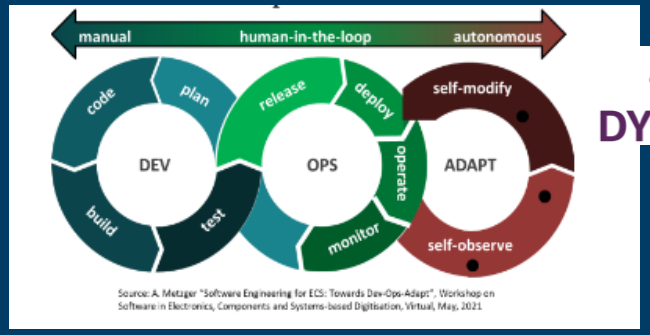
## 1-My (security) research

- About me
- My research roadmap




## 2-SOAR4BC

- (AI-driven) Security Orchestration, Automation and Response for Business Continuity of Critical Infrastructures




DYNABIC

## 3-Lessons learnt

- Explainable AI for Security

SINTEF

# Some recaps and lessons learnt

- SOAR4BC = AI-driven Digital Twin-based Security Orchestration, Automation and Response for Critical Infrastructures
- What is more challenging than developing AI solutions: Knowledge & Tools Unification!
  - Consolidate the contexts from various security aspects, tools, and the digital twin (DT) system
  - Update the “action space” on the fly
  - Support automated tracing and incident analyses
- Automating incident mitigation with ML-based decision-making
- Lessons learnt:
  - AI can support automation but need greater explainability with more context, can be provided by DT.
  - Human-in-the-loop is still required, even though more automation can be applied!
  - Runtime Explainability: Enable security analysts to incorporate domain knowledge in analysing the performance of ML-based security services with comprehensive views into security systems and events.

SINTEF

# Teknologi for et bedre samfunn