ITU Events

ITU Regional Cybersecurity Forum for Europe and CIS

27 – 28 February 2020
Sofia, Bulgaria

Follow us on twitter: @ITU_EUR & @ITUMoscow
www.itu.int/go/EURCIS_CSForum20

ITU Regional Initiative for Europe on enhancing trust and confidence in the use of ICTs

ITU Regional Initiative for CIS on the development and regulation of infocommunication infrastructure to make cities and human settlements inclusive, safe, and resilient

POLICIES & STRATEGIES | CERTIFICATION FRAMEWORKS | DATA PROTECTION | SECURITY CHALLENGES | 5G, eSIM IoT, AI

Hosted and co-organized by:

REPUBLIC OF BULGARIA
State e-Government Agency

REPUBLIC OF BULGARIA
Ministry of Transport, Information Technology and Communications

Cyber for AI <> AI for Cyber

Dr. George Sharkov

Director, European Software Institute CEE & Cyber Security & Resilience Lab (Sofia, Bulgaria) – gesha@esicenter.bg

Adviser Cyber Defense @ MoD, Bulgaria - National Cybersecurity Coordinator (2014-2017) - g.sharkov@mod.bg

Member of EU High Level Expert Group on AI (June 2018 – present, in execution of EU AI Strategy)

- Ethical guidelines for trustworthy AI (EU – High Level Expert Group)
  - Is human oversight always possible/doable?
- Building Technically Robust AI
  - Transparency, Explainability
  - Engineering perspective, requirements
  - “Securing AI” – new emerging standards – ETSI ISG SAI, since October 2019
- Systems-of-Systems & AI
- AI for Red Teaming

EU AI High Level Expert Group - Ethics Guidelines for trustworthy AI (since June 2018)

Human-centric approach: AI as a means, not an end

Trustworthy AI as our foundational ambition, with three components

Lawful AI

Ethically Adherent AI

Technically Robust AI

Three levels of abstraction

from principles (Chapter I)

to requirements (Chapter II)

to assessment list (Chapter III)

Ethics Guidelines for AI – Requirements

- Human agency and oversight
- Technical Robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal & environmental well-being
- Accountability

AI protection and robustness - requirements (from the Assessment List -7 areas)

1. **Human agency and oversight**

- Fundamental rights
- Human agency
- **Human oversight**

2. **Technical robustness and safety**

- Resilience to attack and security
- Fallback plan and general safety
- Accuracy
- Reliability and reproducibility

3. **Privacy and data governance**

- Respect for privacy and data Protection
- Quality and integrity of data
- Access to data

4. **Transparency**

- Traceability
- Explainability
- Communication

Trustworthy AI – the engineering perspective

Quality of AI =

Quality of “knowledge”

+ Quality of Data (learning – ML/DL, use)

+ Quality of technology

+ Quality of software / hardware

+ (Cyber) security

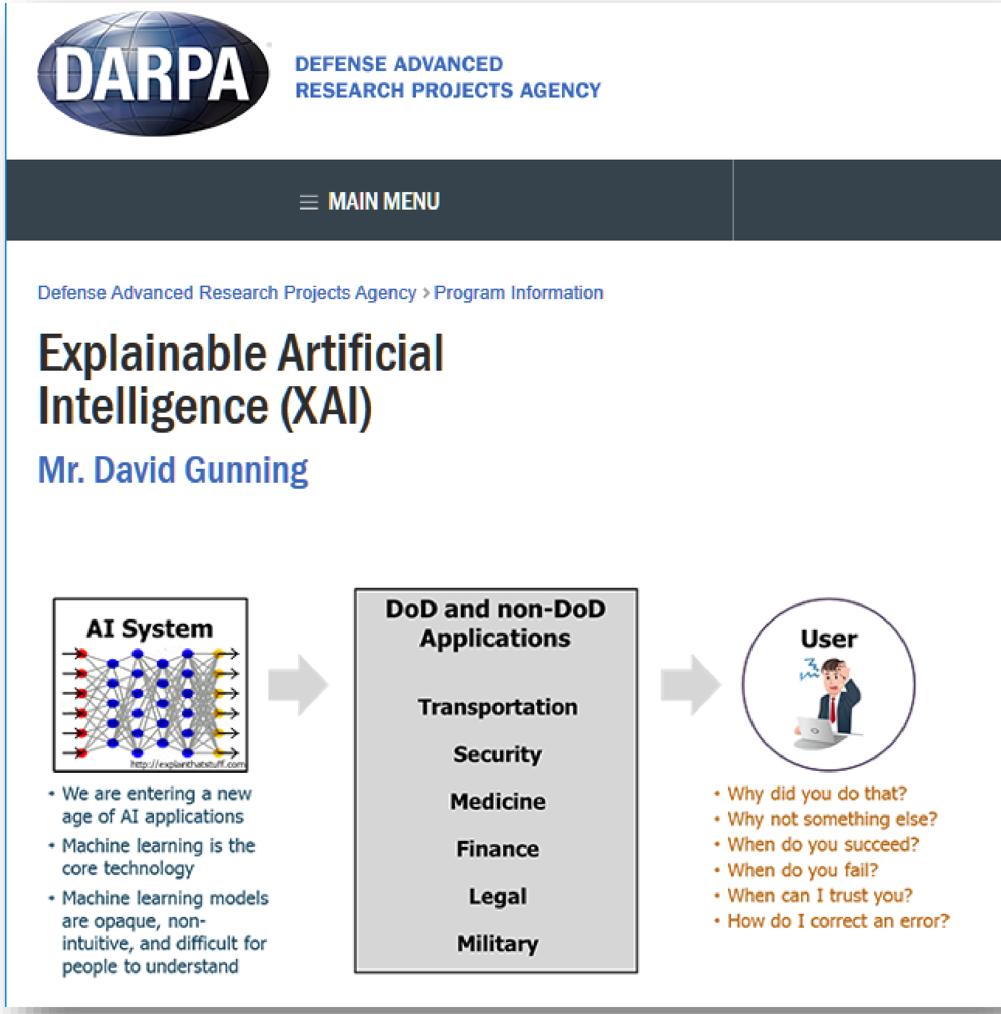
(+ the use in business models and processes – ethical guidelines)

AI systems & safety = “supervising” any ICT / SW systems (e.g. SCADA, ICS)

AI systems and autonomous defense/weapon systems = Explicable/Explainable AI

DARPA program – XAI (Explainable AI)

https://www.darpa.mil/program/explainable-artificial-intelligence



DARPA DEFENSE ADVANCED RESEARCH PROJECTS AGENCY

☰ MAIN MENU

Defense Advanced Research Projects Agency > Program Information

Explainable Artificial Intelligence (XAI)

Mr. David Gunning

AI System

- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand

DoD and non-DoD Applications

Transportation

Security

Medicine

Finance

Legal

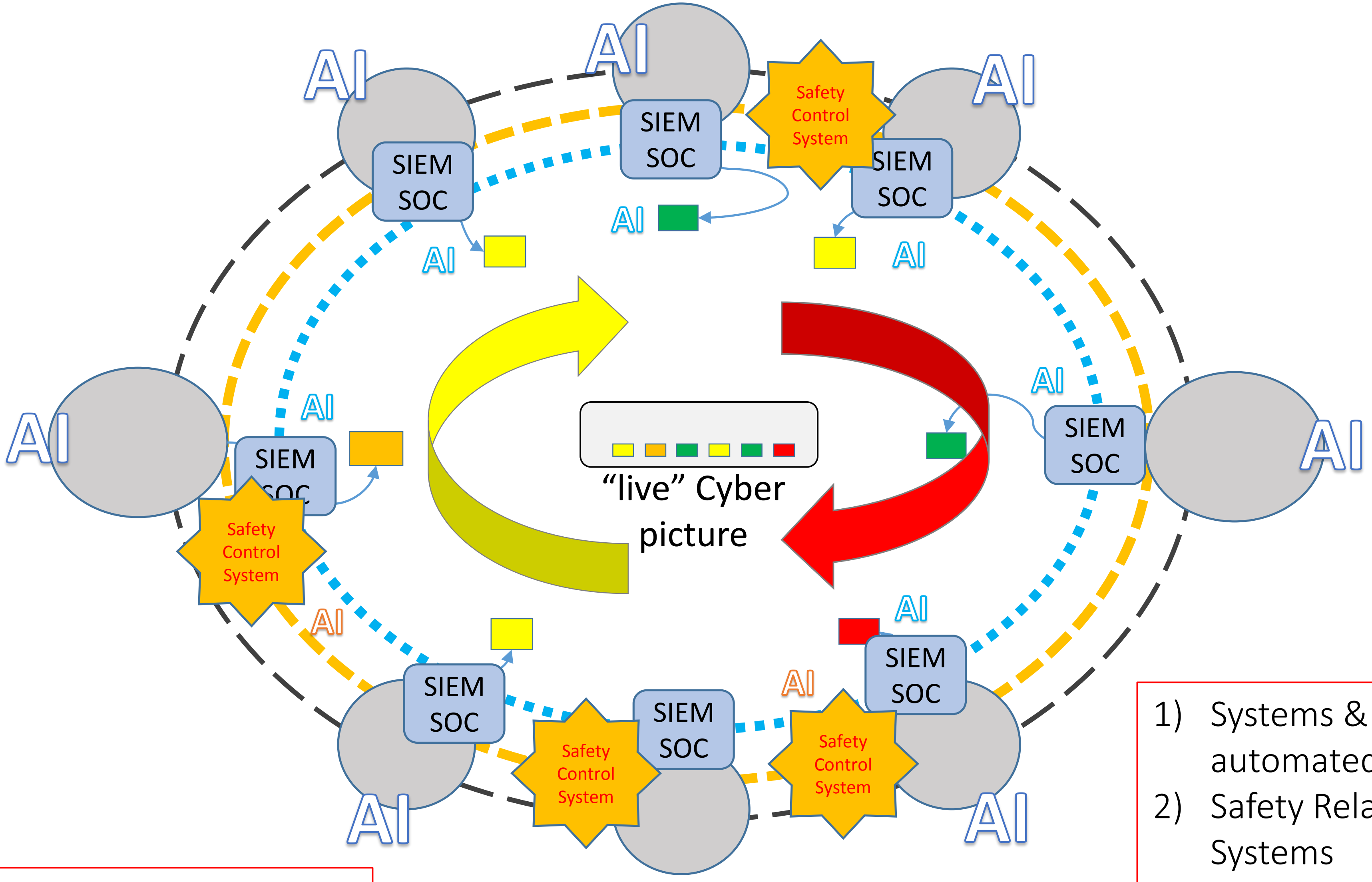
Military

User

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

SoS (Systems-of-Systems) Resilience – new AI collaboration layers:

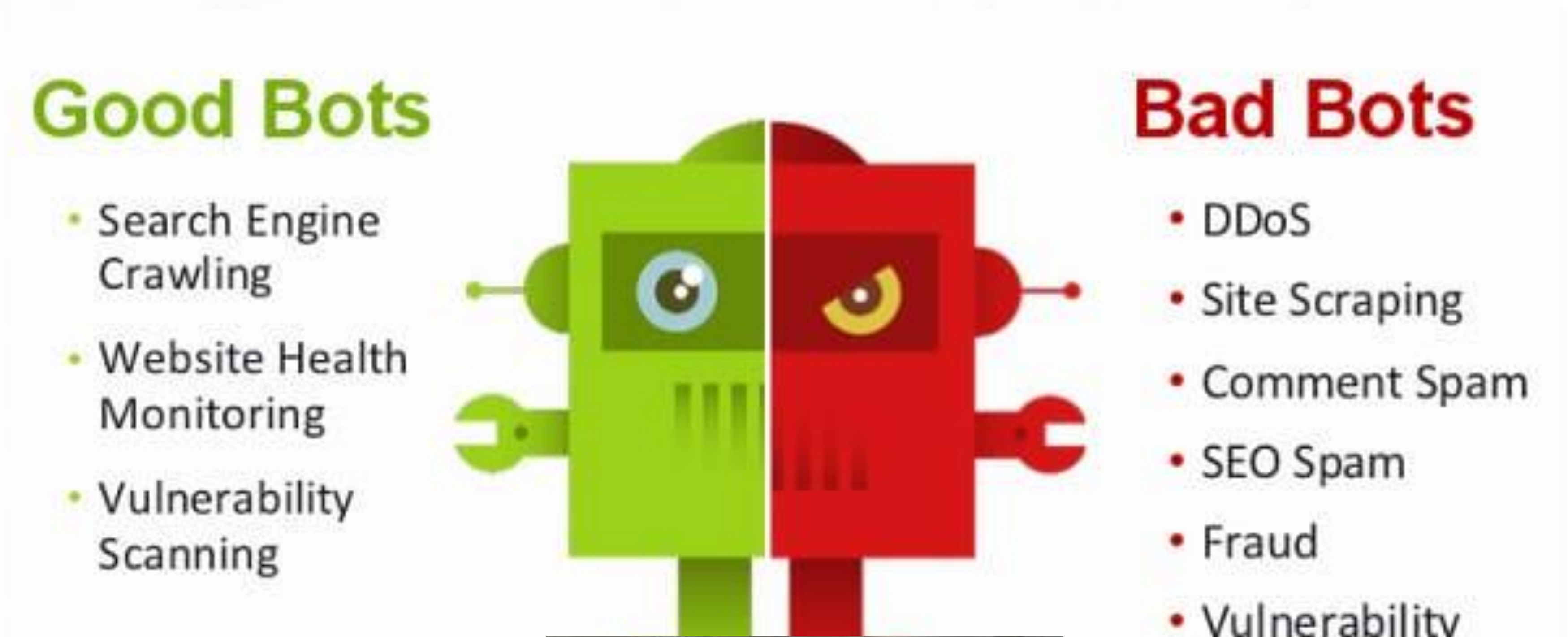
SIEM/SOC collaboration (new layer of SoS) + Safety Systems (another layer of SoS) + … AI (new)



Remember: The "smarter" systems in SoS, The "looser" the coupling is

1) Systems & interoperability – automated, AI
2) Safety Related Control Systems
3) Advanced SIEMs and (A)SOCs:AI/ML empowered

AI vs. AI: Good Bots <> Bad Bots



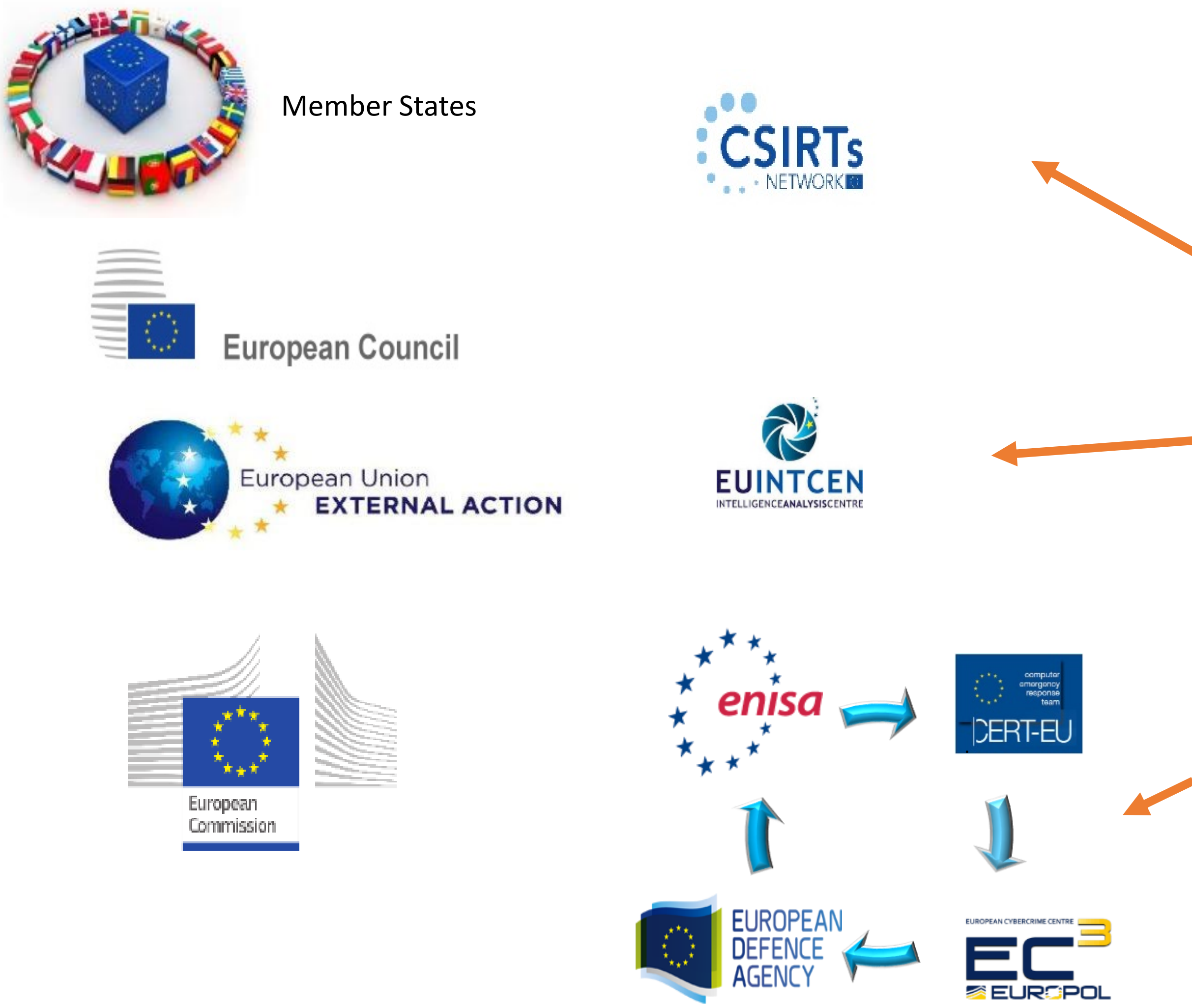
Good Bots

- Search Engine Crawling
- Website Health Monitoring
- Vulnerability Scanning

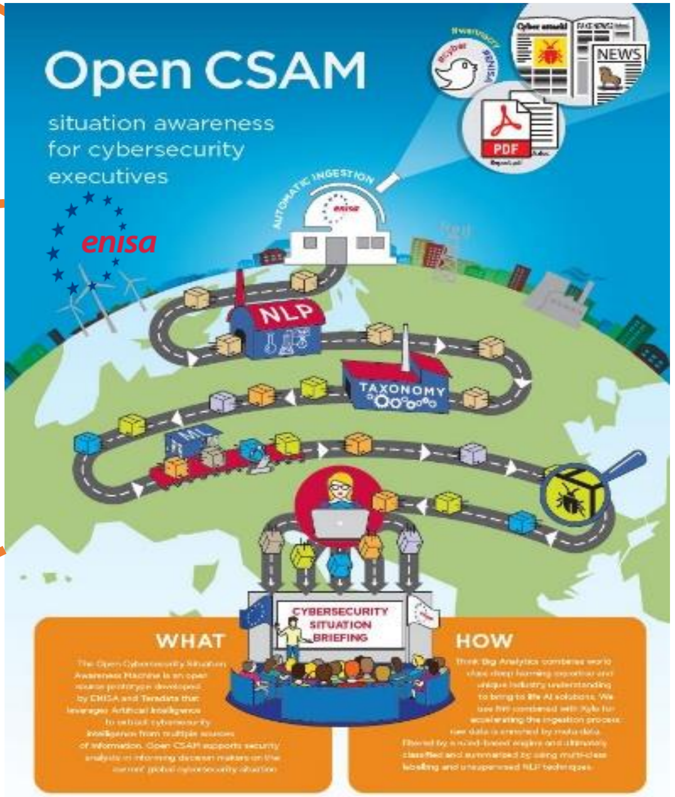
Bad Bots

- DDoS
- Site Scraping
- Comment Spam
- SEO Spam
- Fraud
- Vulnerability scanning

EU: AI support for implementing “large-scale cyber incidents and crises management” (“the Blueprint”, 2017, EU)



An innovative project in support of the Situational Awareness and Incident Response pillars of the Blueprint



WHAT

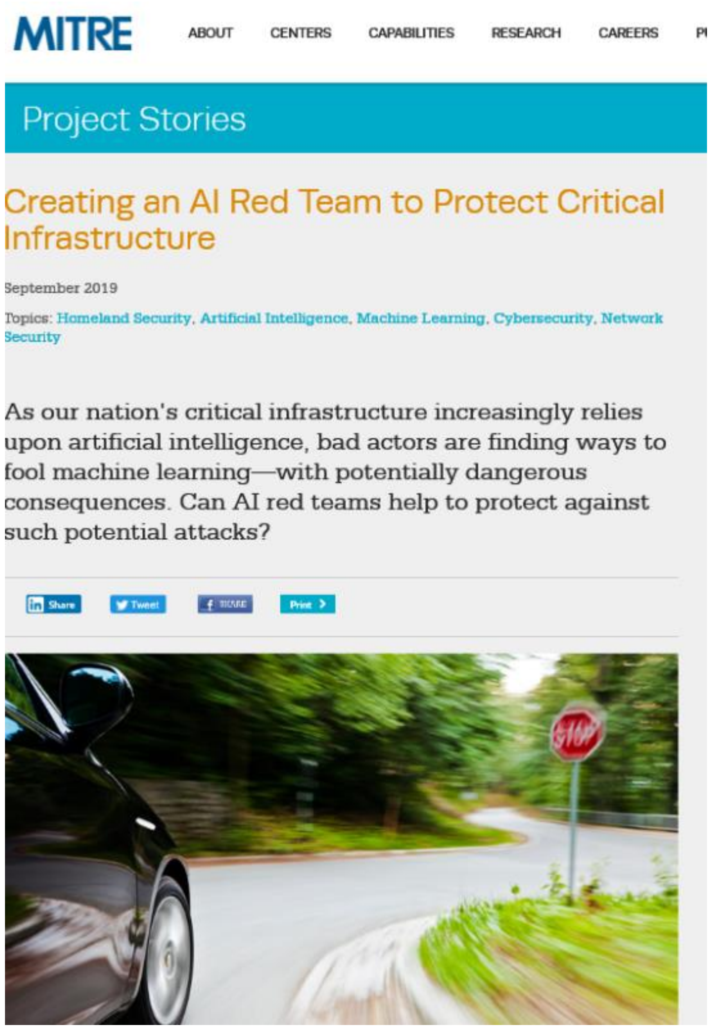
The Open Cybersecurity Situation Awareness Machine is an open source prototype developed by ENISA and Teradata that leverages Artificial Intelligence to extract cybersecurity intelligence from multiple sources of information. Open CSAM supports security analysts in informing decision makers on the current global cybersecurity situation.

HOW

Think Big Analytics combines world class deep learning expertise and unique industry understanding to bring to life AI solutions. We use NiFi combined with Kylo for accelerating the ingestion process: raw data is enriched by meta-data, filtered by a rule-based engine and ultimately classified and summarized by using multi-class labelling and unsupervised NLP techniques.

September, 2019

AI for Read Team



Project Stories

Creating an AI Red Team to Protect Critical Infrastructure

September 2019

Topics: Homeland Security, Artificial Intelligence, Machine Learning, Cybersecurity, Network Security

As our nation's critical infrastructure increasingly relies upon artificial intelligence, bad actors are finding ways to fool machine learning—with potentially dangerous consequences. Can AI red teams help to protect against such potential attacks?

- Operate independently to assure both security and a fresh perspective - keep sensitive data secure, while being transparent about identifying risks
- follow the rapidly evolving landscape of AI attack vectors - what real adversaries know - adversaries may attack anywhere in the ML system lifecycle
- Develop and maintain a counter-AI threat model (ML focused)
- Base recommendations on quantitative evidence - academic metrics that are not always relevant in actual operations (measure both the vulnerability and the potential impact of adversaries attacking real-world systems)

"The good news is that we have the opportunity to start dealing with AI attacks at an earlier stage than we did with cybersecurity"

"The World Wide Web was developed with security as an afterthought, rather than a core design component—and we're still paying the price for it today. With AI, it is not too late to consider safety, security, and privacy before society increasingly relies on this technology."

https://www.mitre.org/publications/project-stories/creating-an-ai-red-team-to-protect-critical-infrastructure

Yes, we did it: BG-GB Cyber Shockwave exercise

March, 2019

Put the "skin in the game" (AI as RED TEAM)

- Industry (*Gas and oil distribution*) >>> State (3 ministries, 3 agencies) interoperability and collaboration (issue)
- Combined Technical + Tabletop (for decision makers):
  4 attack vectors (1 "hidden" on Supply chain)
  + misinformation (web + defacing, fake news/media, mails)
- Small (business) is BIG (threat)
- Context: EU elections (but CYBRID by nature, any time …)

Tested also:

EU Blueprint (ENISA), Cybersecurity Incident Taxonomy,
AI & ML pilot, National legislation/regulations (fiscal system), Standard Operating Procedures (missing or not implemented)

Asymmetry demonstrated:

RED (+simple AI/ML) <> BLUE (Industry + State)

Result: 4 hours, score 3.5 for ??? out of 4

Supported by: UK Embassy, NCSC, UK companies/consultants

What's next (2020): Romania, Greece

ESI European Software Institute
Center Eastern Europe

CY SEC RES LAB

“If you are not part of the solution, you must be part of the problem”

Attributed to: Eldridge Clever (1969); African proverb, others