Trusted IIoT & AI Technologies for the Future of Manufacturing

John Soldatos, Netcompany-Intrasoft
John.Soldatos@Netcompany-Intrasoft.com
Trust in IIoT & AI Systems: The link with the AI Act

- High Quality Data for AI training, testing etc.
- High Robustness & Cybersecurity
- Explainability
- Human Oversight

1. Safety components of regulated products (e.g., medical devices, machinery) which are subject to third-party assessment under the relevant sectoral legislation
2. Certain (stand-alone) AI systems in the following areas:
   - Biometric identification and categorization of natural persons
   - Management and operation of critical infrastructure
   - Education and vocational training
   - Employment and workers management, access to self-employment
   - Access to and enjoyment of essential private services and public services and benefits
   - Law enforcement
   - Migration, asylum and border control management
   - Administration of justice and democratic processes
3. Establish and implement risk management processes & in light of the intended purpose of the AI system:
   - Use high quality training, validation and testing data (relevant, representative, etc.)
   - Establish documentation and design logging features (traceability and availability)
   - Ensure appropriate degree of transparency and provide users with information (on how to use the system)
   - Ensure human oversight (measures built into the system and/or to be implemented by users)
   - Ensure robustness, accuracy and cybersecurity

REMOTE BIOMETRIC IDENTIFICATION (RBI)
Putting on the market of RBI systems (real time and ex-post):
- Ex ante third party conformity assessment
- Enhanced logging requirements
- “Four eyes” principle
  - No additional rules foreseen for the use of real-time and post RBI systems: existing data protection rules apply
Safe, Trusted and Human Centric AI in Manufacturing (Industry 5.0 Outlook)

STAR helps manufacturers and industrial automation vendors to build and deploy Safe Reliable and Trusted Human Centric AI systems

Enable AI systems to acquire knowledge in order to take timely and safe decisions in dynamic and unpredictable environments.

Build AI Manufacturing Systems in-line with Emerging Regulations (AIAct)

H2020 STAR: A Project for Trusted Automation & Artificial Intelligence in Manufacturing

EXPECTED IMPACT

INCREASED INTELLIGENCE & FLEXIBILITY OF PRODUCTION LINES
SAFE HUMAN-ROBOT COLLABORATION AT SCALE
FASTER UPTAKE OF AI SOLUTIONS (QUALITY4.0, CO-BOTS)
ETHICAL IMPACT IN MANUFACTURING IN-LINE WITH HLEG RECOMMENDATIONS
RESEARCH (E.G., SIMULATED REALITY, ACTIVE LEARNING, EXPLAINABLE AI) PLACING EU AT FOREFRONT OF GLOBAL AI R&D
STAR Reference Architecture Model

Cyber Security Domain
- Data Probes & Data Connectors
- AI Cyber Defense Strategies
- Data Provenance & Traceability
- Risk Assessment and Mitigation Engine
- Security Policies Manager

Human Robot Collaboration Domain
- Active Learning
- Production Processes Knowledge Base
- Feedback Module - NLP
- Simulated Realty
- Human Digital Twin

Safety Domain
- Reinforcement Learning
- Mobile Robots Safe Placement
- Human Digital Twin – Worker Safety
- Fatigue Monitoring System
- HMI

Explainable Artificial Intelligence (XAI)

STAR AI / Open Analytics Platform
<table>
<thead>
<tr>
<th><strong>Industrial Data Reliability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environmental influences</strong></td>
</tr>
<tr>
<td>• High or low temperatures, humidity, moisture, and air pressure factors)</td>
</tr>
<tr>
<td><strong>Background noise</strong></td>
</tr>
<tr>
<td>• Noise pollution, interference (alarms, extraneous speech), electrical noise (motors, cooling devices, air conditioning, power supplies)</td>
</tr>
<tr>
<td><strong>Faulty or inaccurate sensors</strong></td>
</tr>
<tr>
<td>• Sensing systems with poor precision.</td>
</tr>
<tr>
<td><strong>Dying battery of a system</strong></td>
</tr>
<tr>
<td>• Compromises its ability to operate properly and provide reliable measurements.</td>
</tr>
<tr>
<td><strong>Compromised or attacked devices</strong></td>
</tr>
<tr>
<td>• Produce biased or fake data due to adversarial attacks (e.g., data modification, false information injection).</td>
</tr>
<tr>
<td><strong>Compromised AI or BigData analytics algorithms</strong></td>
</tr>
<tr>
<td>• Algorithms under poisoning or evasion attacks</td>
</tr>
</tbody>
</table>
Data Reliability and attacks against AI System

Security vulnerabilities coming from AI model errors have become a real concern

State-of-the-art deep neural networks can be easily fooled by a malicious actor and thus made to produce wrong predictions

Explore strategies to generate adversarial examples

Explore Defences Against Adversarial Examples

Goal: Detection mechanism for pinpointing the adversarial examples leveraging Explainable AI
STAR Blockchain Value Propositions

Reliable Algorithms: Provenance of AI algorithms metadata and configurations – “Sealed” algorithms

Reliable AI Outcomes: Provenance of AI analytics outcomes – “Sealed Outcomes”
Defending a Poisoning Attack
From Black-Box AI to Interpretable Models

- Black-box Models (e.g., Deep Learning)
  - Why did you do that?
  - Is there a better option?
  - Is this successful & efficient?
  - Is this a failure?
  - Shall I trust you?
  - When do we get an error?

- XAI Models (e.g., LIME, SHAP etc.)
  - I understand why
  - I understand why there are no better options
  - I know when you succeed
  - I know when you fail
  - I know when I can trust you
  - I know why and when an error occurs
Uses of Explainable AI

1. Explain AI-based decisions to stakeholders (e.g., workers, plant operators)

2. Use the explanation to perform a task e.g.,
   - Analysis: Identify production process configurations that lead to defects - Using Machine Learning / Deep Learning Explainability
   - Autonomy: Decide which tasks can be undertaken by an autonomous system (e.g., drone or robot) - Using Reinforcement Learning Explainability

3. Generating of Credible Synthetic Data - Data Augmentation

4. Identifying Adversarial Actions and Cybersecurity attacks
   - XAI helps signalling abnormal behaviours

5. Legal & Regulatory Compliance
   - Abide by regulatory principles / mandates e.g., transparency, human oversight etc.
   - EU AI Regulatory Compliance

Explaining Quality Inspection – Why is a part defected?

Explanations of classification models

- Image data + Attribution methods
- Produce attribution maps + Visualize into heatmaps
- Highlight features responsible for or against the predicted class

Model-agnostic methods

- Applied to different models
- Produce more general solutions
- Example: LRP variant (local interpretability) + rules

Evaluating the Quality of explanations

- Time complexity -> produce real time results
- Produce human interpretable explanations
Simulated Reality

- Policies learnt in simulation are **safely transferred** to the real world
- **Domain Adaptation** – Shorter round of training in reality to adapt knowledge gained in simulation
- **Domain Randomization** – Produce different simulated training conditions with randomization
- **Randomized-to-Canonical Adaptation Networks (RCANs)** - Convert real world episodes to their simulated equivalent
- **Reliable Data Augmentation**: Addresses the lack of sufficient training data and data skewness (e.g. defective parts much fewer than non-defective)
- **Supervised Learning (e.g. Visual Quality Inspection)**: Synthesis of training samples based on existing ones:
  - Computer Vision (Rotation, Deformation, Noise etc.)
  - Generative Adversarial Networks
  - Variational Auto Encoders
- **Reinforcement Learning (e.g. Part Handling)**:
  - Imitation Learning through robot trajectory logs or human control
  - Reduces amount of trial and error to achieve the task
More Information and Free Download

STAR Web Site: www.star-ai.eu

- Rich Library of Blogs and other publications

Open Access Book (published November 2021; 15500+ Downloads):

Thank you!

Find more:

STAR: https://www.star-ai.eu/
John Soldatos: https://www.linkedin.com/in/johnsoldatos/