SAND2018-11566C

Autonomous Systems, Artificial Intelligence and Safeguards

IAEA Symposium on International Safeguards Building Future Safeguards Capabilities

5-8 November, 2018

PRESENTED BY

Risa Haddal
International Safeguards and Engagements Department

Co-Authors:

Nancy Hayden, Sandia National Laboratories (SNL)

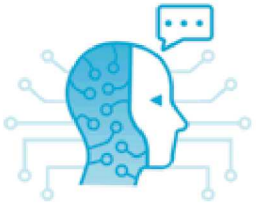
Sarah Frazar, Pacific Northwest National Laboratory (PNNL)

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

PROJECT OVERVIEW

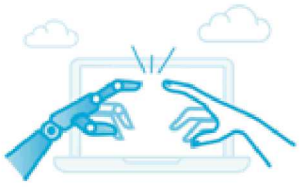
Objective

Explore key safeguards challenges confronting the IAEA today and how autonomous and artificial intelligence (AI) technologies could help.



Artificial Intelligence

Machine Learning



Human-Computer Interaction

Chat Bot

Framework

Framework for identifying and evaluating autonomous and AI methods consists of five elements:

1) I.D. principal safeguards verification challenges;
2) Develop criteria that autonomous and AI methods would need to address;
3) Develop inventory of methods;
4) Develop safeguards use cases;
5) Technical Evaluations.

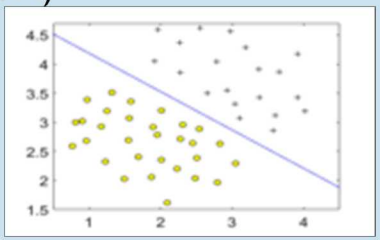
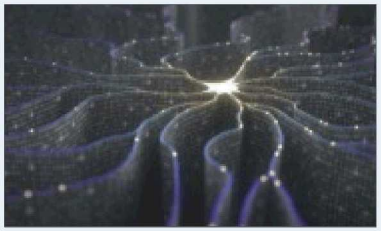
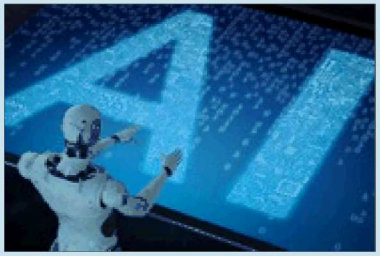
1. SAFEGUARDS TECHNICAL CHALLENGES

Safeguards Challenges			
Increase in new types of facilities and next generation reactors.	Increase in number of facilities under safeguards. 	Global nuclear expansion of trade in equipment, materials and know-how.	Increasing data flows from information collection systems. 
Need to protect safeguards information.	Transmit secure, authentic communications. 	Increased amount of nuclear material under safeguards.	Need for efficient and effective technology acceptance/adoption.
Increases in environmental samples, sample processing and data management. 	More countries bringing into force the Additional Protocol.	More states with Broader Conclusions.	Constrained IAEA resources. 

2. CRITERIA

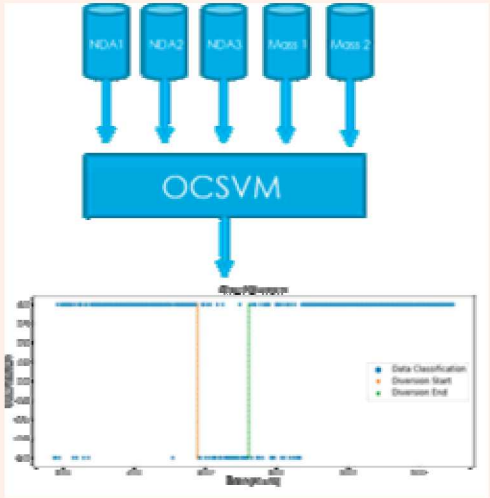
Safeguards Challenges					
		Increase in new types of facilities; next-gen reactors	Increased # of facilities under safeguards	Consistent need to protect safeguards information and transmit secure communications	Increasing data flows from safeguards data collection systems
Technical Objective: Verify that Declared Materials Remain in Peaceful Uses	Activity: Perform inspections; audit nuclear material inventories; perform PIV	Criteria: 1) Reliably identify and explain significant anomalies 2) Reduce time in the field without reducing quality of SGs inspections and conclusions; 3) Verification of Operator declarations; 4) Increase efficiency and productivity while reducing costs.	Criteria: 1) Reliably identify and explain significant anomalies); 2) Reduce time in the field without reducing quality of SGs inspections and conclusions; 3) Verification of Operator declarations; 4) Increase efficiency and productivity while reducing costs.	Criteria: 1) Reliably control access to and ensure security of information and communications; 2) detect threats to information security; 3) Reliable, secure and controlled access to computer networks; 4) detection of threats to computer networks; 5) backup systems for operating through attacks on computer networks.	Criteria: 1) Reliably identify and explain significant anomalies without violating SGs agreements (IP, legal constraints, data authentication); 2) Sufficient, resilient bandwidth for data flows; 3) timely detection and adjustment to capacity limitations; 4) verify normal operations; 5) identify outliers and anomalies; 6) detect meaningful patterns across time and space for SGs conclusions; 7) Increase efficiency and productivity while reducing costs.

SELECTED METHODS

	METHOD	DESCRIPTION
1.	One-Class Support Vector Machine (OCSVM) 	• Traditional support vector machines (SVM) trained in a one-class sense • ML models with associated learning algorithms for classification and regression analysis • An unsupervised SVM that is trained on data that can be divided into classes
2.	Convolutional Neural Networks (CNN) 	• A class of deep, feed-forward (information moves in only one direction) artificial neural networks (computing systems that learn tasks based on examples, e.g. image recognition) • Successfully applied to analyzing visual imagery • Applications in image and video recognition, recommender systems and natural language processing
3.	Networked Autonomous Robots 	• Networked system of cooperative, interactive, autonomous robots managed through AI • Leverages cloud processing of observations and environmental data • ML behavior-based system allows adaptive tasking of robotic units • System tasks robotic devices connected via a wired and/or wireless communication network

USE CASE 1A

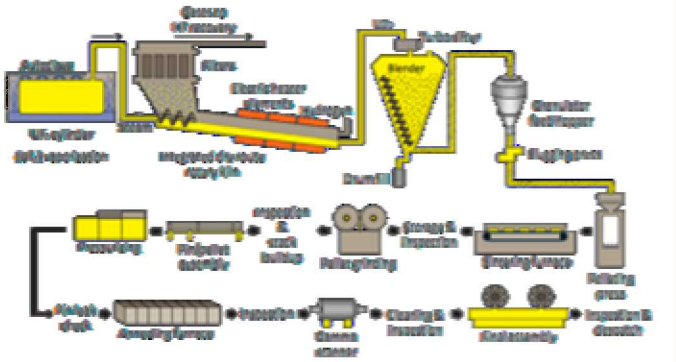
Use Case Name	OCSVM for Data Analysis and Anomaly Detection at Reprocessing Plants
Description	Use case explores reducing amount of person days in the field, by using unattended monitoring data at a reprocessing plant to train an OCSVM algorithm to detect anomalies or off-normal conditions.
Basic Flow	1. Unattended monitoring and sampling data from reprocessing plant from a period of normal operations is used to train an OCSVM algorithm. 2. OCSVM is used to automatically monitor unattended data in an MBA where Pu and U are separated. 3. During automated review, OCSVM indicates a threshold has been reached for detection of a possible anomaly, triggering onsite inspection. 4. Anomaly + results of onsite sampling leads to determination that there have been other minor differences in solution inventories, indicating possible protracted diversion. 5. Further inspection and discussion required with senior inspector and plant operator.



Graphical representation of a safeguards system using the OCSVM.
Source: Shoman, N., Cipiti, B. Unsupervised Machine Learning for Nuclear Safeguards. INMM Annual Meeting, 2018.

USE CASE 1B

Use Case Name	Convolutional Neural Networks (CNNs) for Physical Inventory Verification
Description	A CNN is used to support PIV at a fuel fabrication plant. An IAEA inspector reviews slow-frame video from the Next Generation Surveillance System (NGSS). Inspector focuses on fuel rod assembly hall over the past year to verify the number of assemblies as declared by operator to IAEA in support of material accountancy. CNN is trained to recognize standard PWR fuel assemblies using thousands of open source and IAEA archived images. Slow-frame video data is uploaded onto trained CNN software program. After filtering, CNN software identifies fuel assemblies, clusters or classifies them, and calculates an item count number. Result indicates that the number of assemblies in the hall matches the number declared, reducing or eliminating the need for an inspector to do a physical count.



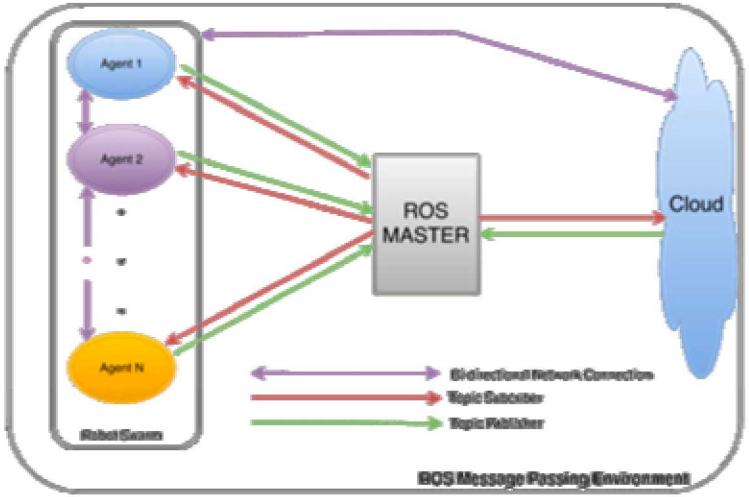
Source: *International Safeguards in the Design of Fuel Fabrication Plants*, IAEA Nuclear Energy Series No. NF-T-4.7. IAEA. Vienna, Austria. 2017.

1. Inspector downloads CNN software on laptop at IAEA HQ, prepares for onsite inspection;
2. Inspector arrives at fuel fabrication plant; accesses slow-feed video (1 frame/second);
3. Surveillance data uploaded onto laptop and run on CNN software.
4. Trained CNN software filters through imagery.
5. Once filtering is complete, CNN software clusters fuel assemblies into objects and counts total number of assemblies to support PIV.
6. CNN software produces a time series of counts identifying that there are 64 of 193 fuel assemblies in the hall awaiting packaging and shipment.
7. IAEA inspector compares this number against number of assemblies declared by the operator and verifies that they match, eliminating need to do a physical count.
8. Inspector records findings and writes a report.

USE CASE 1C

Use Case Name	Networked Autonomous Robotics to Improve Safeguards at Encapsulation Plants
Description	A networked autonomous robotics system managed through cloud processing is used to support safeguards activities at an encapsulation plant, including inventory counting of canisters in temporary storage. Robots collect data through on-board cameras and sensors for object recognition and counting, environmental monitoring, and event detection. Data is transmitted to the cloud where ML methods such as anomaly detection or Bayesian processing are deployed for adaptive control of the robots.

Basic Flow



Robotics Operating System.
Source: Miratabzadeh, Seyed et al. Cloud Robotics: A Software Architecture. 2016 World Automation Congress. IEEE. Rio Grande, Puerto Rico. 06 October 2016.

1. Cloud network comes online; autonomous Robot units deployed at encapsulation facilities.
2. Network pings robot units for signs of life; Allocates first instructions;
3. Nuclear fuel arrives at encapsulation plant.
4. Network anticipates or prompts for verification data;
5. Robot unit locates canister ID or markers and sends ID data to network;
6. Network compares data with operator declarations and other information in database. If consistent, networks confirms verification. If inconsistent, networks identifies discrepancy, self-checks with backjumping, remotely transmits report to inspectorate and operator.
7. Robot locates canisters and IDs in temporary storage; sends ID and location data to network.
8. Network confirms consistency of ID and location in data archive

TECHNICAL EVALUATION

Use Case 1A: OCSVM for Data Analysis and Anomaly Detection at Reprocessing Plants			
Safeguards Deployment Options	Benefits	Risks and Challenges	Timeframe
Anomaly detection at bulk handling or reprocessing facilities with large amounts of heterogeneous unattended monitoring data.	Save time and resources for analysts and inspectors by detecting anomalies without requiring labeled data; With appropriate historical data on normal and off-normal conditions, OCSVM proof-of-concept could be easily demonstrated to build trust and transparency.	Collection of training data could be expensive depending on required size of data set. Robust feasibility study or proof-of-concept needed. Corruption or manipulation of training data or input data by adversary might be feasible.	5 years to test/validate. Deployment thereafter, depending on IAEA safeguards acceptance process.

TECHNICAL EVALUATIONS CONTINUED

Use Case 1B: CNNS for PIV			
Safeguards Deployment Options	Benefits	Risks and Challenges	Timeframe
Image recognition and item counting to support PIV.	Reduce time in the field for item counting. Provide more nuanced picture of activity in a facility that a human may not otherwise see with naked eye. Detect anomalies or off-normal conditions (Note: Depends on quality of training data.)	Improper training or poor quality data could result in false alarms or ability to recognize items. Training data could be vulnerable to manipulation. Explaining how CNN makes decisions is open area of research. Verified testing needed to facilitate acceptance and deployment.	5 years to test and validate approach, possibly longer depending on IAEA safeguards acceptance process.

TECHNICAL EVALUATIONS CONTINUED

Use Case 1C: Networked Autonomous Robotics to Improve Safeguards at Encapsulation Plants			
Safeguards Deployment Options	Benefits	Risks and Challenges	Timeframe
Support safeguards verification activities such as PIV (inventory count) and material verification. Autonomous robots execute physical tasks; cloud environment centralizes data processing, supports anomaly detection, and facilitates information management.	Enable more frequent inventory verification; Reduce potential vulnerabilities such as injury or mistakes due to human behavioral factors (e.g., fatigue, inattention); Increased accuracy and verifiability of assessments due to substantially more data being collected and recorded; Strong cost/benefit (assuming accurate performance); and Immediate distribution of measurement results to multiple locations;	Cyber vulnerabilities of cloud network; False alarm rates might be high due to challenging environments and sparse training data; System behavior might not always be intuitive or explainable, raising questions about trust; Difficult terrain or environments limit robot mobility or sensing; Manual verification of autonomous system findings, clearing false alarms, and maintenance could exceed workloads for manual verification	1-3 years for deployment of pilot system. 10 years or more for full implementation.

CONCLUSION

- Autonomous systems, AI and ML provide important opportunities to improve effectiveness and efficiency of IAEA safeguards.
- Current and emerging methods hold potential for:
  - Reducing time and resources needed for data analytics, inspections and safeguards implementation;
  - Recognizing patterns or anomalies that might not otherwise be observed by humans.
- Non-trivial challenges:
  - Training and quality of data will impact systems' ability to learn, decide and act;
  - Manual re-verification could reduce efficiency; and
  - Training data and networks could have vulnerabilities.
- Bottom Line: Robust training and evaluation required to increase trust, transparency, and likelihood of adoption.

QUESTIONS?