

Explainable AI in RAMS

Jan 24-27, 2022

RAMS 2022, The 68th Annual Reliability &
Maintainability Symposium, Tucson, AZ

Navid Zaman, MEng

January 2022



Overview & Outline

Slide

- Introduction 3.
- Defining Explainable AI 4.
- Applications in RAMS 5.
- Metrics and Tests for Certification 6.
- Domain Insights 7.
- Use of Insights 8.
- Summary 9.

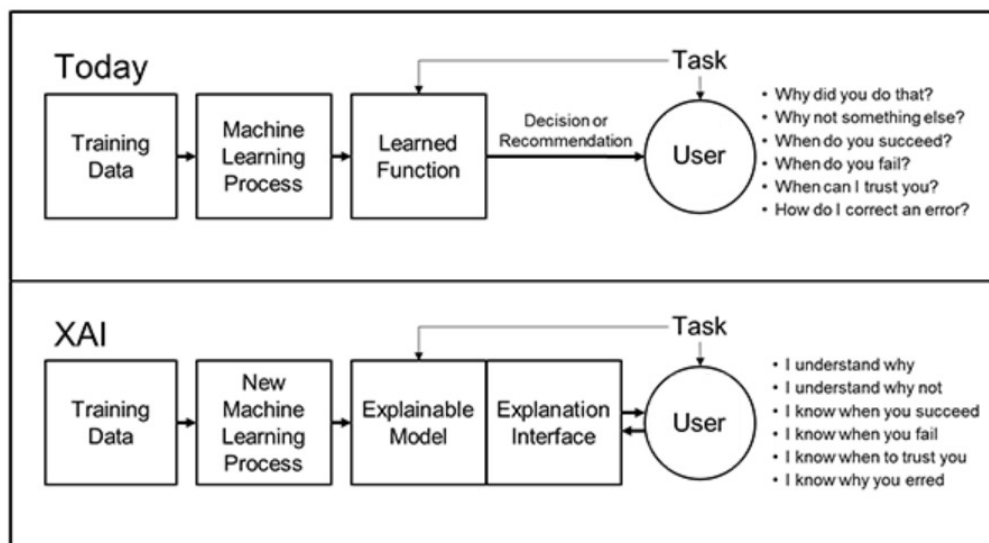
Introduction

- Artificial Intelligence (AI) gives engineers the ability to identify failures, manage useful life and optimize maintenance to better improve Mission Assurance of a system.
- Understanding or clarity in the AI output begins to be challenged.
- Today's AI provides users with outputs without much context around the prediction.
- Explainable AI or XAI aims to resolve the issue of a 'black box' in Machine Learning (ML).



Definition of Explainable AI

- XAI is an AI which the results of a solution/prediction can be understood by the user.
- Explainable AI or XAI aims to solve:
 - Reasoning around why AI made a certain decision
 - Reassurance to the user that the AI output is correct
 - Removal of 'black box' ML and introduction of 'white box' ML
 - Build trust between user and AI



Applications in RAMS

- **Fault Detection and Isolation**
 - Concerned with monitoring of an asset for the purpose of identifying failures/faults and locating where in the system the failure/fault has occurred
 - Read sensor data and understand what constitutes a normal reading compared to a failed state.
 - Uses classification and novelty algorithms
- **Maintenance Optimization**
 - The ability to enhance the planning and scheduling of maintenance tasks to better im-prove organization/asset performance goals such as availability and cost of maintenance
 - Works with FDI for determination of maintenance tasks to perform
 - Discrete optimization algorithms and goal-based strategies are used
- **Residual Useful Life Estimation**
 - The determination of an asset's remaining time that it can perform its intended function
 - The main technology utilized is Deep Learning based regression
 - Common ideas circulate the use of confidence bounds, heuristics and constraints to read results more cautiously
- **Guided Reporting and Inspection Tool**
 - The use of a predefined taxonomy for maintenance and failure reporting into a Maintenance Management System (MMS) for consistent and trace-able reporting
 - AI techniques such as Natural Language Processing (NLP) and image classification are used to better enhance the reporting capability

Metrics and Tests for Certification

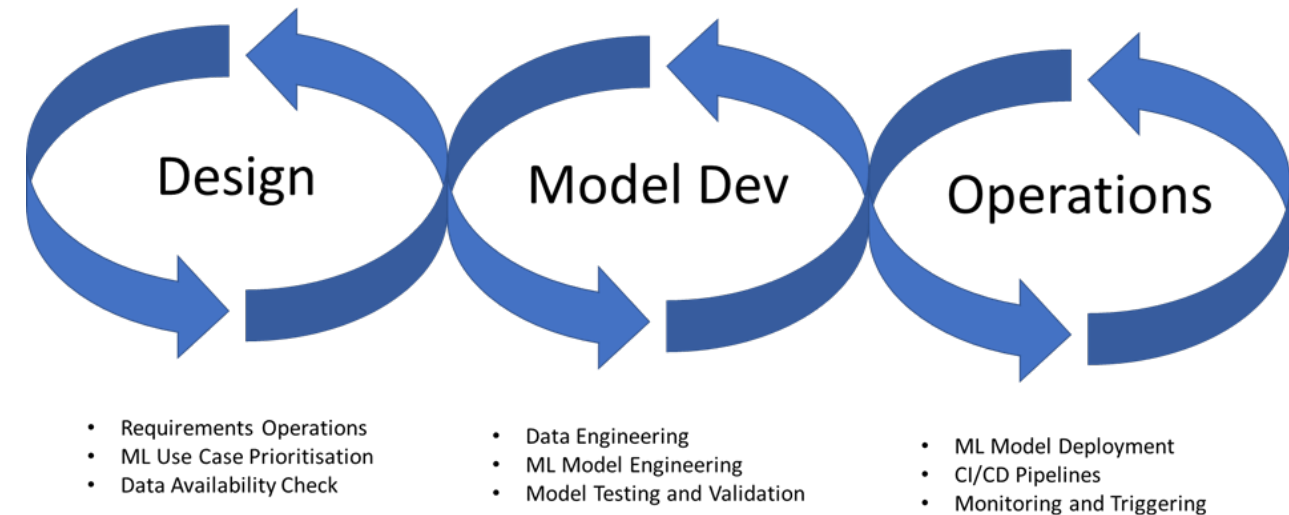
- Transparency during inference is just as important as that observed during training and testing
- Assessment and training is crucial to establish trust
- Confusion Matrices and Derivatives:
 - Classification of multilabel and binary sets
 - Accuracies and F1 scores
- Regression Metrics:
 - Forecasting of time series signals
 - MSE, MAE and RMSE etc.

Domain Insights

- **Information which would be useful to show as part of the XAI output.**
- **Syndromes and Diagnostic Rules**
 - The syndrome of failure is a unique combination of sensor responses which point to a particular failure of an asset in a system
 - Diagnostic rules provide the user with the expected combination of sensor responses to correctly identify the failure in a descriptive format
- **Sensor Information**
 - Probability of Detection (PoD)
 - Specificity
 - Sensor Type
 - Confidence Intervals
- **Loading Conditions and Operating Modes**
 - Operational Loading Conditions
 - Environmental Loading Conditions
 - Operating Modes - different modes in which a system and its assets are expected to operate for a given time period

Use of Insights

- Automated Deployments: ML-Ops
 - Using unique and displayed triggers for ML model deployment
 - Improving metric management
- Data Fusion
 - Virtual sensing – calculation of important variables using sensor data
 - Improved operating context awareness for failure determination



Chris Jefferson, "Machine Learning: Automated Devops and Threat Identification", 2020

Summary

- The industry is quickly adopting the use of AI in its tools and processes but are often faced with the issue of black-box solutions
- Explainable-AI or XAI has become a popular and growing field of study - how to make results from these technologies more accessible.
- This paper has listed and discussed various strategies to extract and expose useful information beside the predictions that can be utilized in meta-analysis, interpretations and backtracking.
- Further applications of such observations are also proposed to take full advantage of AI in RAMS.

A black and white photograph of an F-35 fighter jet in flight, viewed from a low angle. The jet is flying towards the left, with its wings and canards visible. The cockpit is clearly visible, showing the pilot. The background is a bright, cloudy sky. The image is split vertically: the left half is in color, and the right half is a dark, semi-transparent overlay.

PHM Technology

Decisions better made

Thank you