Global Power System Transformation Research Topic 3 – Control Room of the Future

2022-23 Stage 2 Research - Final Report

Commonwealth Scientific and Industrial Research Organisation (CSIRO)
23 June 2023

Contents

Executive Summary

The Global Power System Transformation (G-PST) research roadmap, developed by CSIRO and EPRI in 2021, for the Control Room of the Future required advancements in the field of artificial intelligence (AI) and machine learning (ML). The control room is potentially advantageous for machine learning applications, given the proliferation of time series data and the need for new automated solutions. With this in mind the stage 2 developments focussed on the key, early-stage enablers for development of machine learning in system operations.

This report summarises the work carried out by EPRI as project lead with Royal Melbourne Institute of Technology (RMIT) and Australian Electricity Market Operator (AEMO) in stage 2. The report details the nine key challenges for advancement of AI in the energy sector, with particular focus on system operations.

Given the difficulties in developing AI/ML applications, a methodology was developed which could be used to assess the efficacy and risks inherent with the potential development of use cases. The methodology and framework build on both the intelligent smart grid use case methodology and the United States of America National Institute of Standards and Technology (NIST) AI risk management framework.

A long list of possible use cases was collated and shortened to a list of potentially advantageous use cases for control room and real time operations. Alarm management and Event Analysis was identified as the most appropriate use case for development within this project. The selected use cases were analysed and documented using the framework and risk assessment.

The project team worked together to develop algorithms based on a synthetic dataset of 5.8 million alarms that were then tested on real world alarm data from AEMO.

Four approaches to the use case objective were identified:

- **Alarm Forecasting** – to attempt to predict short term future alarm loads based on input and historical data. Machine learning approaches were used for the forecast, with promising early-stage results.
- **Alarm Spike Detection** – to identify ‘spikes’ in alarm loads, relative to medium term baselines and to group the alarms within the spikes into analysable buckets. Statistical and machine learning approaches were used for this, both of which showed promising results.
- **Alarm Chatter Detection** – to identify if alarms were ‘chattering’ i.e., spurious continuous alarming of single points that distract attention from operators. Machine Learning was used for this with good results.
- **Pattern Matching** – To identify patterns of features within the alarm spikes such as words. Machine learning was used for this, but results were inconclusive and further research work is needed/

The project was successfully completed and a framework for the development of the main alarm management use case and future use cases was established. Given the difficulty of establishing AI/ML in operational environments, the work completed as part of this project will be an important enabler of future progress in this field and allows stage 3 research work to be continued in the years ahead. The aim is to develop robust AI/ML algorithms for use in operations control rooms in Australia and worldwide, to streamline the process of operating decarbonised energy networks. Developing these capabilities advance the CROF research roadmap for the data pillar by approximately 10%, but the key work in this stage was in developing the enabling capabilities for future AI / ML use cases for control rooms.

Acknowledgement

“The research presented in this report was funded by CSIRO, Australia’s national science agency, and carried out as part of CSIRO’s contribution to the initiatives of the Global Power System Transformation (GPST) Consortium. This research supports Australia's transition to a stable, secure and affordable power system and contributes to critical research identified by the Consortium required to accelerate the decarbonisation of our electricity grid. More details can be found at ***https://www.csiro.au/en/research/technology-space/energy/G-PST-Research-Roadmap*** **[csiro.au]**.”

1. Introduction

1.1 Project Background and Context

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) have launched a research program to develop the research priorities identified by the Global Power System Transformation (G-PST) collaborative. EPRI (Electric Power Research Institute) are leading the team to conduct research on the Topic 3 – the Control Room of the Future (CROF).

The first stage of the work was completed in 2021[1] and consisted of a research roadmap outlining pathways for innovation around five key pillars:

- IT Architecture,
- Data,
- Control room tools and applications,
- Human factors and
- Facilities.

The roadmap developed a comprehensive five stage pathway for each of the key pillars, to evolve towards the Australian transmission control room of the future by 2030.

The research continued in 2022 with a second stage of the research to begin development of the tasks on the CROF research roadmap. Given the broad scope of the roadmap, the stage 2 CROF project, documented in this report, narrowed the focus to tasks that require significant initial development work. The focus of stage 2 was on the early stages of the data and control centre tools pillars.

1.2 Relationship to the CROF Research Roadmap

The primary aim of the project is to develop the capability in the artificial intelligence (AI) and machine learning (ML) fields for real time power system applications, with particular focus on developing use cases for AI/ML in the operational context.

This topic area was chosen for advancement in Stage 2 because of the long gestation period of applications of this nature, due to:

- The complexity of machine learning model and algorithm development and iteration.
- The risks and trust elements of deploying in high-reliability environments
- The need for large quantities of data to train and test any models developed.
- The need for significant computing power, for training, labelling and simulation.
- Availability and accessibility of real or realistic data datasets for the training of AIML algorithms and models.

AI/ML research and projects, by their nature, are experimental and at an early stage for system operations and control room applications. However, experimental code and software is not acceptable for high reliability systems like system operations control centres, with major safety and economic risks if something goes wrong. This is even more acute for AI/ML given the inherent lack of understanding about how really advanced AIML systems that use neural networks work. Traditionally developed software tools are developed by software developers iteratively with code and continuous testing. AIML approaches are

[1] CSIRO Research Plan for Topic 3 Control Room of the Future Link: **https://www.csiro.au/-/media/EF/Files/GPST-Roadmap/Topic-3-Control-Room-of-Future-Final-Report-with-alt-text-2.pdf**

generally not developed in this way. This requires the need for a clean, risk-based methodology to be in place before proceeding with development. I

AI/ML development is an iterative process and not something one can or should "plug in" to highly complex Operational Technology (OT) system architectures. Its development needs foundational work with data management, computer power, training, testing, iterations, trust, tuning etc. For this reason, data governance, management and standardisation for alarm and operational data were early stage enabling milestones on the data roadmap (see Figure 1). In the CROF research roadmap for the control room tools, there was a clear need identified for a new solution for alarm, asset health and disturbance investigation tool, as shown (in Table 1) these were the use cases most applicable for development in stage 2 of the project.

In addition to the CSIRO CROF research roadmap, the operations technology roadmap developed with CSIRO and AEMO in 2022 [2] identified AI/ML developments as part of a continuous activity across the decade, for all the operational capabilities, see Figure 2 (boxed in yellow).

By aligning with both the CROF 2021 research roadmap for CSIRO and the Operational Technology Roadmap for AEMO and selecting a use case that is applicable to all control rooms in the energy sector – the work completed in stage 2 aligns with the research roadmap and puts in place strong foundations for future developments and applications.
The research roadmap for the data models and streaming pillars are shown in Figure 1, with indicative dates for the milestones. For clarity, the stages indicated in this roadmap do not correspond to the yearly stages of the annual CSIRO research cycles.

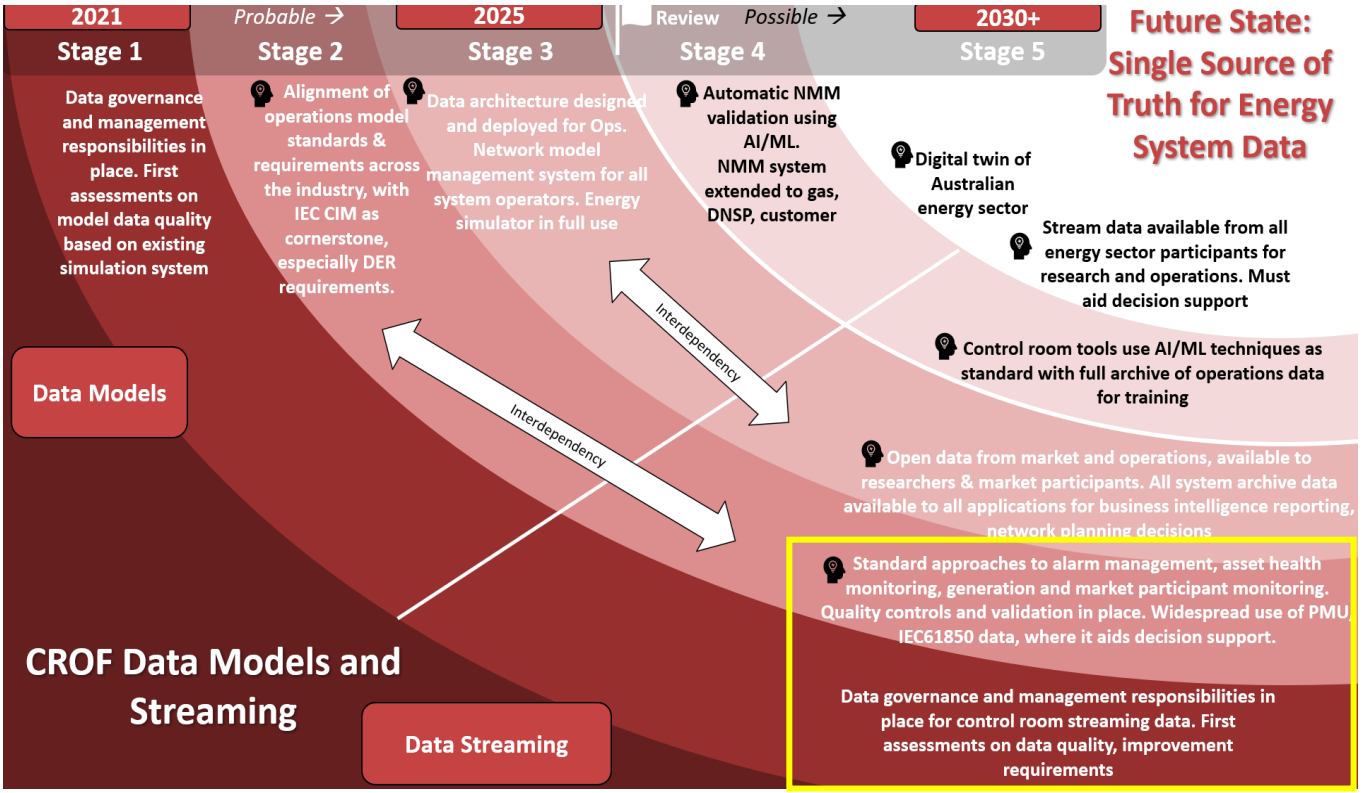


Figure 1 CROF roadmap for data models and streaming from the CSIRO Stage 1 roadmap. Yellow box indicates the areas of focus for the research work in 2022-23 for the stage 2 of the project.

[2] AEMO Operations Technology Roadmap 2022 Link: **https://aemo.com.au/-/media/files/initiatives/operations-technology-roadmap/executive-summary-report-for-the-otr.pdf**

Table 1 One of the control room tools and applications identified for development in the CSIRO CROF Stage 1 Research Roadmap

Control Room or Operations Planning and Support Tool	Inputs	Outputs	Ideal Future State
Asset Health Alarm Root Cause Analysis / Disturbance Investigation Tool	SCADA alarm data, asset health data, previous disturbances and logs, weather data, protection data	Visually presents concise message indicating root cause of disturbance with mitigation.	Umbrella tool for all alarm data that instantly identifies alarm & disturbance root cause and directs operator's attention to issues and solutions. Likely to use ML.

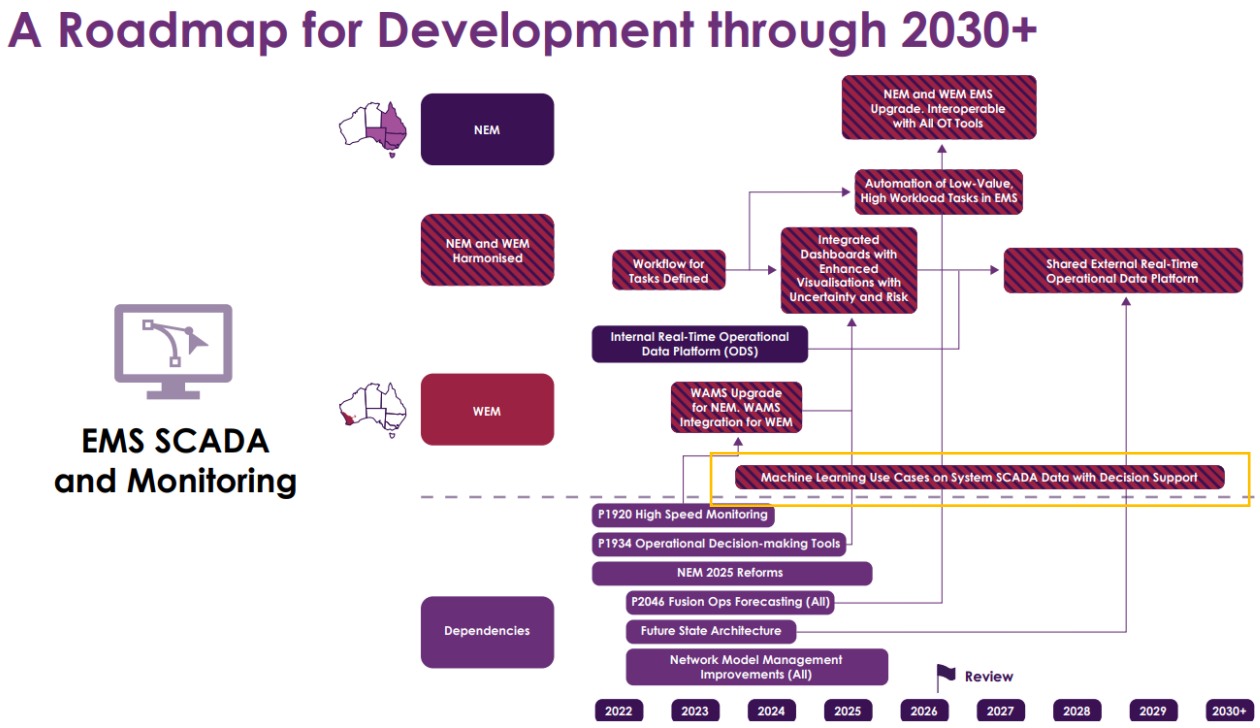


Figure 2 AEMO CSIRO Operations Technology Roadmap 2022 - showing AI/ML developments with SCADA data and decision support as a continuous process through the decade, aligning with the CSIRO CROF research roadmap.

1.3 CSIRO Research Roadmap Progress

The stage 2 work developed some of the early stage enabling activities of the data pillar of the research roadmap. The indicative progress is shown below in Table 2. Progress towards these milestones will be progressed further in future stages in the years ahead.

Table 2 Indicative progress of the milestones of the CROF research roadmap for the data pillar

	RESEARCH TASK	Completion progress
	CROF Data Models and Streaming	
1	Data governance and management responsibilities. First assessments on model quality based on existing simulation system.	10 %
2	Alignment of operations model standards& requirements across industry	5 %
3	Data Governance and management responsibilities for control room streaming data.	*10 %*
4	Standard approaches to alarm management, asset health monitoring, generation and market participant monitoring.	30 %
5	Open data from market and operations, availability.	5 %

1.4 Project Team

The stage 2 CROF project team was made up EPRI, Royal Melbourne Institute of Technology (RMIT), AEMO with valued input from CSIRO, listed in Table 3. The team worked collaboratively and the EPRI are thankful for the valued contributions from all.

Table 3 Project team for CROF Stage 2

Company	Name	Role
EPRI	Adrian Kelly	Project Manager
	Mobolaji Bello	Assistant Project manager
RMIT	Prof. Xinghuo Yu	Machine Learning RMIT Team Lead and Machine Learning Subject Matter Expert
	Dr. Chen Liu	Machine Learning Subject Matter Expert and Developer
	Geordie Dalzell	Machine Learning Subject Matter Expert and Developer
AEMO	Karin Rodrigues	System Operations Subject Matter Expert
	Tjaart van Der Walt	System Operations Subject Matter Expert
	Luke Robinson	System Operations Subject Matter Expert
	Elena Kranz	Data Scientist
CSIRO	Mahathir Almashor	Machine Learning Subject Matter Expert
	Thomas Brinsmead	CSIRO Project Manager

2. Background Research and Landscape Assessment

2.1 The Challenges with Deployment of Artificial Intelligence and Machine Learning

While artificial intelligence and machine learning (AI/ML) are in focus across all aspects of society, useable use cases in the system operations and control room context are very limited and are challenging to develop. Some of the challenges with developing and deploying AI/ML use cases in operations and the control rooms of the world are shown in Figure 3 and detailed below.

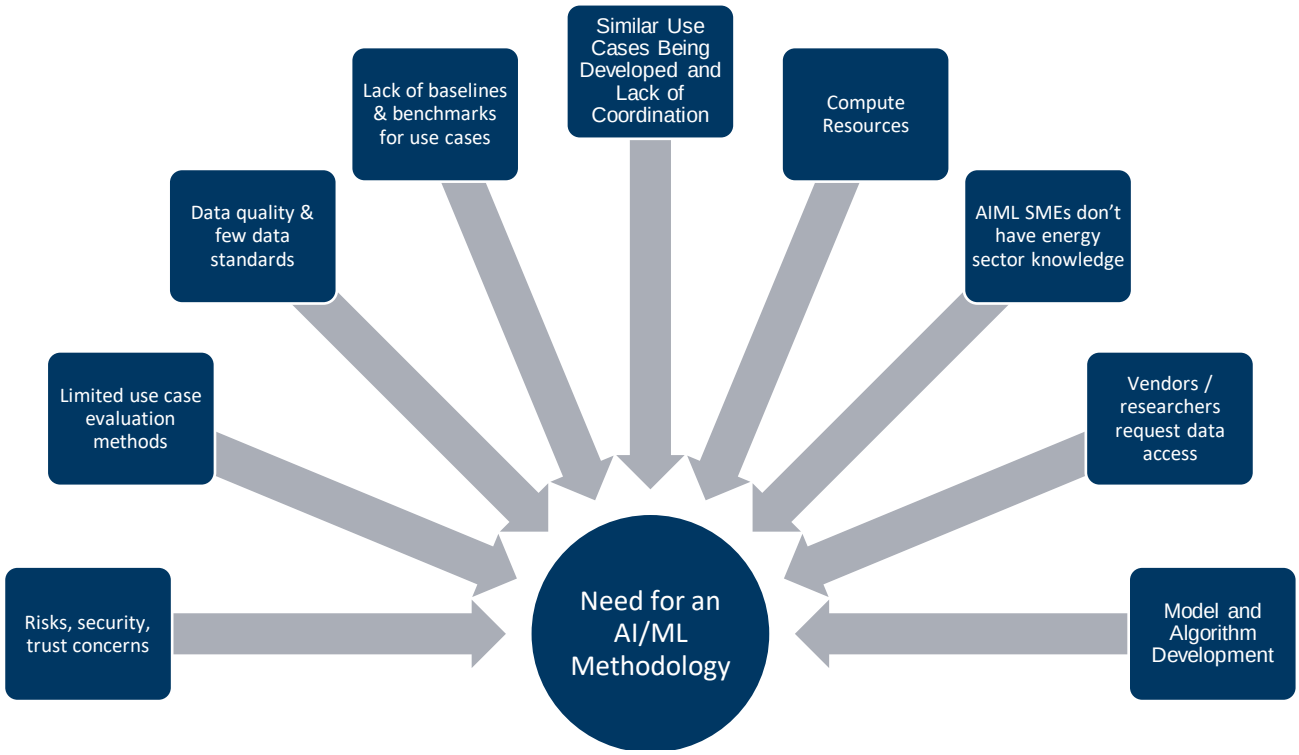


Figure 3 Challenges with developing machine learning use cases for transmission network operations

2.1.1 Challenge 1 – Risks, Security and Trust Concerns

With AI/ML that use deep neural networks (the most frequently used models today), there is an element of "black box" development, where an output is derived from inputs plus weights - without an explanation for how the output was achieved. In context with less associated criticality, this approach may be acceptable. In the control room and power and market systems operations contexts, trust in algorithms is very important, given the wider socio-economic risks to the electricity network if a wrong decision is made. Security of the algorithms and wider systemic risk are also important to AI/ML development given the context.

Potential Controls

- Select lower criticality use cases for initial development. For example, developing models and algorithms for safety switching (with health and safety impacts) rather than time series analysis and forecast use cases.
- Adopt the key elements of the NIST (United States National Institute of Standards and Technology) AI Risk Management Framework into the AI/ML Development Methodology.

2.1.2 Challenge 2 – Use Case Evaluation Methods

Machine learning is only applicable to a limited set of use cases, given the data inputs and required outputs. Machine learning has proven effective for time series data use cases (such as forecasting) language pattern recognition, and image recognition. For example, optimisation techniques have been used for decades in power system analysis and have proven successful, despite compute time limitations. Machine learning may not be a relevant upgrade or replacement to standard optimisation in some cases so a methodology for evaluating the use cases relevance to potential machine learning should be developed.

Potential Controls:

- ML Use Case Methodology with evaluation parameters and reference to existing solutions in the industry.

2.1.3 Challenge 3 – Data Quality and Limited Data Standards

Machine learning on time series data is generally only useful with large quantities of high quality, structured data sets. The datasets in typical utilities are from multiple vendors with bespoke database structure and disparate quality of the datasets that need to be rationalised, combined and cleaned before they are useful as inputs to an algorithm.

Potential Controls:

- For each use case develop standards for the data inputs, leveraging existing industry data standards, such as Common Information Model (CIM).
- Use synthetic datasets as a proof of concept to develop the most appropriate data input standard.

2.1.4 Challenge 4 – Lack of Baselines and Benchmarks

A baseline in machine learning parlance corresponds to the existing performance of a system, from which a newly applied algorithm (such as a machine learning application) can improve upon. This can be established as the baseline from which future algorithms can be measured, to determine performance. In many cases, machine learning is applied without knowledge of what the baseline performance is. Sometime this is difficult to quantify, but an effort should be made to attempt to evaluate the current performance. Benchmarks follow from baselines and are the current performance of an algorithm, that algorithms should strive to surpass with tuning and improvements.

An example might be in alarm management: Baseline is 500 alarms per hour. An algorithm should strive to reduce that while maintaining operational information and limiting computational time.

Potential Controls:

- For each use case establish a baseline performance and within the methodology keep track of benchmarks in algorithmic performance.

2.1.5 Challenge 5 – Similar Use Cases Being Developed and Lack of Coordination

Data science teams in electricity system utilities will tend to work on similar use cases as neighboring system operators, without collaboration, despite similar data and available technology. For example, asset health analytics use case will use similar data and algorithms in every network location e.g., transformer gassing levels, temperature. Demand forecasting is another relevant example which takes similar inputs and outputs. In some cases, data scientists work independently from the engineering subject matter experts or the wider electricity and power system communities.

Utilities in Australia face similar challenges to utilities across the world. These challenges include DER integration, forecasting, congestion management, market operations. There are few forums for knowledge

share between utilities for AIML applications at the global level and not much in terms of direct collaborations. This has led to a lack of major innovations in the field. The European Union has recognised this issue and is at an early stage of engagement on research projects to develop digital twins and AI experimentation test facilities[3]. The UK is developing the "virtual energy system"[4] concept to achieve similar outcomes. China have made advances in utilizing AI for energy sector use cases in recent decades. However, global collaboration network is limited.

Potential Controls

- Develop a global collaborative forum for AI/ML developments in the energy sector, utilising elements of the use case development methodology.
- Identify data science teams through collaborative networks and share information via forums such as conferences, webcasts and workshops. Leverage and combine resources across entities to develop solutions that work across the industry.
- Develop and share list of use cases for the energy sector with maturity levels, baselines, benchmarks with contacts and all relevant information and knowledge associated with the use case.

2.1.6 Challenge 6 – Compute Resources

Effectively developed machine learning usually requires vast computing resources to train models on massive datasets with environments to develop and hone models and algorithms. Cloud based services are the most frequently used approaches to the challenges of compute resources in the industry. However, because of data sensitivity and security issues with cloud-based web services they are not favoured at present in the energy sector. On-prem high performance computing capability is cost prohibitive without obvious value and benefit. This is slowly changing, with the use of private cloud and regulators considering the value of cloud solutions over the limitations of on prem solutions.

Potential Controls

- Develop a framework for using cloud-based services securely for specific use cases and ML development for the electricity sector.

2.1.7 Challenge 7 – AI/ML Data Science SMEs lacking Energy Sector Knowledge

For some complex use cases domain specific knowledge is very important to algorithm development. This is especially true in the system operations and control room contexts. Given the data quality challenges and the complexity of the processes that are managed in the control room it is difficult to transfer the knowledge from the real world to the development environment. Labelling of datasets is also very important in some applications, which requires SMEs and generally is difficult to accomplish without prior experience.

Potential Controls

- Involvement of domain experts in use case development, labelling and testing.

2.1.8 Challenge 8 - Vendors and Research Data Access

There are many large software vendors and software as a service (SaaS) provider for cloud computing that can leverage their existing systems and platforms to develop use cases and add value to operational processes. In addition, there are many innovative early stage and start-up software vendors with limited background in the industry and energy sector that are requesting data access to trial their applications and platforms. Finally, the research and academic community around the world are working to develop applications and have disparate connections to utilities within their country.

[3] EU Horizon Proposal for an AI Testing Experimentation Facility in 2024 Link: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl5-2024-d3-01-11

[4] National Grid Electricity System Operator Virtual Energy System Link: **https://www.nationalgrideso.com/future-energy/virtual-energy-system**

Potential Controls

- Develop a testbed with synthetic datasets for testing by vendors and the academic community.
- Develop secure ways to share anonymized archive datasets that the vendor and research academic community can utilise to innovate and develop applications, with reference to benchmarks.

2.1.9 Challenge 9 – Model and Algorithm Development

The quality and sustainability of open-source libraries and code for development of AIML models and algorithms is a challenge. In many cases machine learning is an exploratory topic with experimentation, iteration, testing and improvement over time. Open-source tools are used by developers in the industry such as Tensor Flow, Pytorch and OpenAI Gym. Development practices, frameworks, and standards should be streamlined across all use cases in development. Beyond proof-of-concept prototypes, there should be a clear development cycle to ensure software is sustainably developed and easily deployable in production environments.

Potential Controls

- Develop and maintain a repository of safe and secure libraries, code, packages and open-source tools for use in development.
- Maintain a framework for model development and testing that includes coding standards, commenting, documentation, user interface development and the software life cycle.

2.2 Summary

The challenges identified are significant barriers to AIML development and innovation in the system operations context. The project aims of a methodology and use case development though this methodology will allow for strong foundations for future AI/ML use cases and for sustainable development of AIML in the control room context going forward.

3. CSIRO CROF Stage 2 Project Plan

3.1 Research Project Objectives

The risks of deploying machine learning and AI use applications in operations can be mitigated by a methodology for safely developing the research and applications. There are methodologies for AI/ML development, but none are specific to the field of transmission operations or the electricity sector more broadly.

To advance the field of artificial intelligence and machine learning development in the electricity system operations domain, a new methodological approach to machine learning projects is required. In this project, two key objectives are developed:

A. Develop a use case methodology, risk assessment and data assessment for the applying of AI/ML to system operations and control room developments.

B. Develop a test use case with the methodology using synthetic data and real data (from AEMO) to prove or refine the efficacy of the methodology.

The methodology was developed and tested with the development of a use case by machine learning specialists from RMIT.

3.2 Research Relevance to Australia

Currently there is no existing machine learning project methodology for electricity or energy system use case development. In addition, despite widespread industry adoption, there are limited machine learning applications in the electricity control room and power system operations domains, or the energy sector more broadly, both in Australia and around the world. This is mostly to do with the challenges identified in section 2.1, primarily trust concerns but also a lack of demonstrable value based on existing applications. Operations works with systems where innovation is difficult and existing systems are preferred to experimental solutions.

This project aims to add structure to the development of machine learning applications so that they are baselined and benchmarked and can be applied by AEMO and other transmission network operators. The development of the methodology and use case in this project can be used by researchers and practitioners in Australia, however, a secondary aim is to also to generalise and extend the methodology and framework, to be applicable to industries beyond energy and beyond Australia.

3.3 Research Approach

The Machine Learning Use Case Methodology (MLUCM) was developed with insight from all project members. The initial set of use cases was collated and evaluated by the project team, with a small number of candidates use cases to be selected for advancement and development by the project team.

4. Development of an AI/ML Use Case Methodology

4.1 Machine Learning Use Case Methodology (MLUCM)

The machine learning use case methodology (MLUCM) is a sustainable, risk-based approach to developing machine learning use cases in the control room operations context, to address the challenges detailed in Section 2.

The proposed methodology includes both basic background information and technical aspects such as the eight broad topic areas, shown in Figure 4 and detailed in Section 4.5.

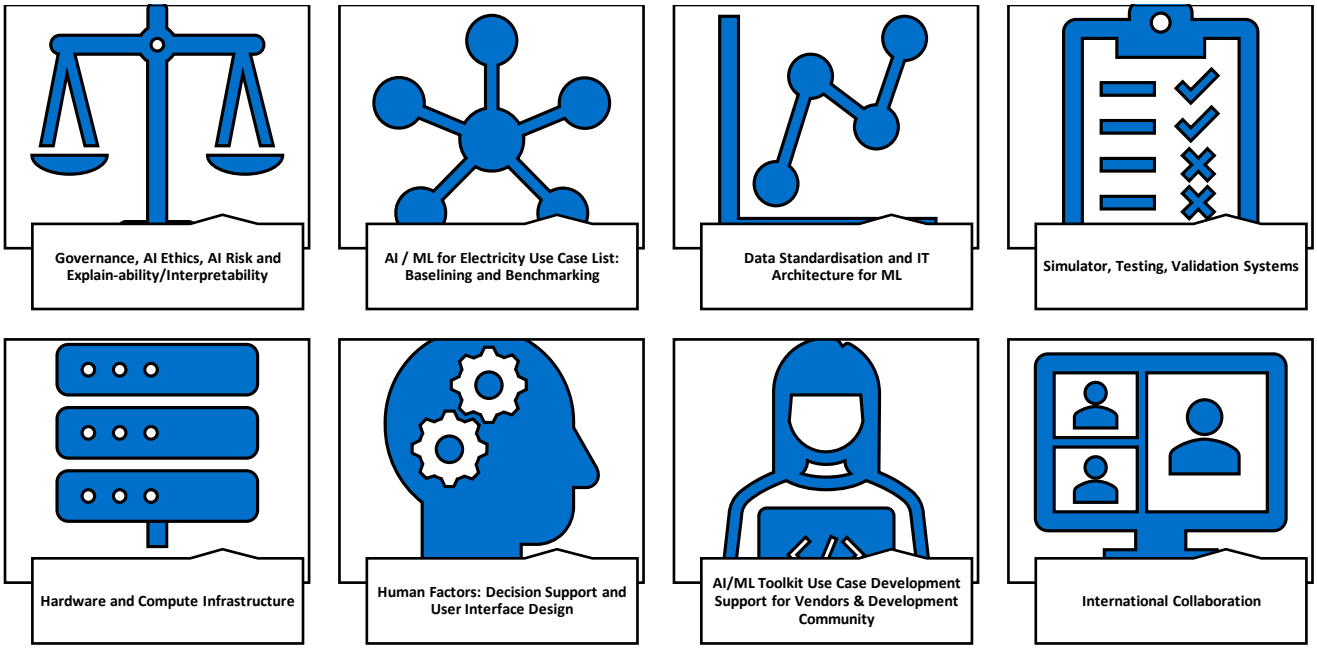


Figure 4 Proposed topic areas of the ML use case methodology for electricity use cases with focus on control centre applications

4.2 EPRI and IEC Use Case Methodology

In the early 2000s EPRI developed the Intelligrid Smart Grid Methodology process for developing energy sector smart grid applications and use cases. Once this research and development process reached its goal it was released to the research and development community through the International Electrotechnical Council (IEC) standards body as IEC 62559[5].

While not a direct AI/ML application use case methodology, given its origin for energy sector and smart gird application, this methodology - and variations of it to address the unique challenges of AI/ML applications - for use case development is chosen as the most appropriate starting point for development but is adapted to address the challenges of developing AI/ML use cases and the ML project life cycle.

[5] IEC 62559 Use Case Methodology Link: **https://syc-se.iec.ch/deliveries/iec-62559-use-cases/**

4.3 Introduction

4.3.1 Name of Use Case

This section details basic background information

Use Case Identification		
ID	***Domain(s)***	***Name of Use Case***

4.3.2 Version Management

This section details basic version control management information

Version Management					
Version Management Changes / Version	***Date***	***Name Author(s) or Committee***	***Domain Expert***	***Area of Expertise / Domain / Role***	***Approval Status draft, for comments, for voting, final***

4.3.3 Scope and Objectives of Use Case

This section outlines the objective and the scope of the use case

Related business case	
Objective	
Scope	

4.3.4 Narrative of Use Case

This section has a more complete description of the use case

Narrative of Use Case
Short description – max 3 sentences

4.3.5 General Remarks

If there are any general remarks, they can be included in this section

General Remarks

4.4 Diagrams of Use Case

Diagrams helpful to the use case description can be drawn here. This may not be necessary in all use cases.

Diagram of Use Case

4.5 Technical Details

4.5.1 Risks, Security and Trust

What are the risks and how will it be managed? The National Institute of Standards and Technology (NIST) AI Risk Management framework is the world leading AI risk management framework [6]. It has broad applicability to energy sector use cases and can be applied as a test case for the MLUCM. Not all points will be applicable for all use cases, but comments can be added to explain the context specific information.

Risk Framework: NIST AI Risk Management Framework V2.0		
Map	MAP 1.1: Intended purpose, prospective settings in which the AI system will be deployed, the specific set or types of users along with their expectations and impacts of system use are understood and documented. Assumptions and related limitations about AI system purpose and use are enumerated, documented, and tied to (Test Evaluation, Verification, Validation) TEVV considerations and system metrics.	
	MAP 1.2: Inter-disciplinary AI actors, competencies, skills, and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized.	
	MAP 1.3: The business value or context of business use has been clearly defined or – in the case of assessing existing AI systems – re-evaluated.	
	MAP 1.4: The organization's mission and relevant goals for the AI technology are understood.	
	MAP 1.5: Organizational risk tolerances are determined.	
	MAP 1.6: Practices and personnel for design activities enable regular engagement with stakeholders and integrate actionable user and community feedback about unanticipated negative impacts.	
	MAP 1.7: System requirements (e.g., "the system shall respect the privacy of its users") are elicited and understood from stakeholders. Design decisions take socio-technical implications into account to address AI risks.	
	MAP 2.1: The specific task, and methods used to implement the task, that the AI system will support is defined (e.g., classifiers, generative models, recommenders).	
	MAP 2.2: Information is documented about the system's knowledge limits and how output will be utilized and overseen by humans.	
	MAP 2.3: Scientific integrity and TEVV considerations are identified and documented, including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability), and construct validation.	
	MAP 3.1: Benefits of intended system functionality and performance are examined and documented.	

[6] NIST AI Risk Management Framework Link: https://www.nist.gov/itl/ai-risk-management-framework

Risk Framework: NIST AI Risk Management Framework V2.0		
	MAP 3.2: Potential costs, including non-monetary costs, which result from expected or realized errors or system performance are examined and documented.	
	MAP 3.3: Targeted application scope is specified, narrowed, and documented based on established context and AI system classification.	
	MAP 4.1: Approaches for mapping third-party technology risks are in place and documented.	
	MAP 4.2: Internal risk controls for third-party technology risks are in place and documented.	
	MAP 5.1: Potential positive and negative impacts to individuals, groups, communities, organizations, and society are regularly identified and documented.	
	MAP 5.2: Likelihood and magnitude of each identified impact based on expected use, past uses of AI systems in similar contexts, public incident reports, stakeholder feedback, or other data are identified and documented.	
	MAP 5.3: Assessments of benefits versus impacts are based on analyses of impact, magnitude, and likelihood of risk.	
Measure	MEASURE 1.1: Approaches and metrics for quantitative or qualitative measurement of the most significant risks, identified by the outcome of the Map function, including context-relevant measures of trustworthiness are identified and selected for implementation. The risks or trustworthiness characteristics that will not be measured are properly documented.	
	MEASURE 1.2: Appropriateness of metrics and effectiveness of existing controls is regularly assessed and updated.	
	MEASURE 1.3: Internal experts who did not serve as front-line developers for the system and/or independent assessors are involved in regular assessments and updates. Domain experts, users, and external stakeholders and affected communities are consulted in support of assessments.	
	MEASURE 2.1: Test sets, metrics, and details about the tools used during test, evaluation, validation, and verification (TEVV) are documented.	
	MEASURE 2.2: Evaluations involving human subjects comply with human subject protection requirements; and human subjects or datasets are representative of the intended population.	
	MEASURE 2.3: System performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions like deployment setting(s). Measures are documented.	
	MEASURE 2.4: Deployed product is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.	
	MEASURE 2.5: AI system is evaluated regularly for safety. Deployed product is demonstrated to be safe and can fail safely and gracefully if it is made to operate beyond its knowledge limits. Safety metrics implicate system reliability and robustness, real-time monitoring, and response times for AI system failures.	
	MEASURE 2.6: Computational bias is evaluated regularly, and results are documented.	
	MEASURE 2.7: AI system resilience and security is evaluated regularly and documented.	
	MEASURE 2.8: AI model is explained, validated, and documented. AI system output is interpreted within its context and to inform responsible use and governance.	
	MEASURE 2.9: Privacy risk of the AI system is examined regularly and documented	
	MEASURE 2.10: Environmental impact and sustainability of model training and management activities are assessed and documented	

Risk Framework: NIST AI Risk Management Framework V2.0		
	MEASURE 3.1: Approaches, personnel, and documentation are in place to regularly identify and track existing and emergent risks based on factors such as intended and actual performance in deployed contexts.	
	MEASURE 3.2: Risk tracking approaches are considered for settings where risks are difficult to assess using currently available measurement techniques or are not yet available.	
	MEASURE 4.1: Measurement approaches for identifying risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. Approaches are documented.	
	MEASURE 4.2: Measurement results regarding system trustworthiness in deployment context(s) are informed by domain expert and other stakeholder feedback to validate whether the system is performing consistently as intended. Results are documented.	
	MEASURE 4.3: Measurable performance improvements (e.g., participatory methods) based on stakeholder consultations are identified and documented.	
Manage	MANAGE 1.1: Determination is made about whether the AI system achieves its intended purpose and stated objectives and should proceed in development or deployment.	
	MANAGE 1.2: Treatment of documented risks is prioritized based on impact, likelihood, and available resources methods.	
	MANAGE 1.3: Responses to the most significant risks, identified by the Map function, are developed, planned, and documented. Risk response options can include mitigating, transferring, sharing, avoiding, or accepting.	
	MANAGE 2.1: Resources required to manage risks are considered, along with viable alternative systems, approaches, or methods, and related reduction in severity of impact or likelihood of each potential action.	
	MANAGE 2.2: Mechanisms are in place and applied to sustain the value of deployed AI systems.	
	MANAGE 2.3: Mechanisms are in place and applied to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.	
	MANAGE 3.1: Risks from third-party resources are regularly monitored, and risk controls are applied and documented.	
	MANAGE 4.1: Post-deployment system monitoring plans are implemented, including mechanisms for capturing and evaluating user and stakeholder feedback, appeal and override, decommissioning, incident response, and change management.	
	MANAGE 4.2: Measurable continuous improvement activities are integrated into system updates and include regular stakeholder engagement	

4.5.2 Equivalent Use Cases

Details of how do other equivalent entities solve the problem, requiring use case development?

Equivalent Use Cases	
Neighbouring utility / regional level stakeholder use case application	
Country level stakeholder equivalent applications	
Global level stakeholder equivalent applications	
Academic and research review	
Opportunities for collaboration with stakeholders locally, regionally	
Academic and research collaboration explored?	

4.5.3 Use Case Evaluation Methods

Details as to why AIML is being considered for this use case.

Evaluation	
Why is current practice insufficient/inefficient?	
Why is machine learning being considered?	
What alternatives were considered and why were they discounted?	

4.5.4 Data Standards

What data is being used in the use case, what is the data quality and is there a data standard?

Data Standards	
What data is required?	
Data Source (Database)	
Data type (Time Series/Image/Text)	
Data Standard (CIM, 61850.)	
Data security levels (critical energy, personal data)	
Structured / Unstructured data	
Data Quality (Low/Medium/High)	
Data Cleaning Time	
Data Housing (Location – on shore, on shore, cloud etc)	
Potential Inclusion of PII (Personal Identifiable Information)	
Assessment of Potential for Data Biases in Models	

4.5.5 Baselines and Benchmarks

How is success measured and evaluated over time and what is the latest status.

Baselines and Benchmarks	
What metric will be used for evaluation model performance (speed, accuracy)?	
How will the baseline metric be measured?	
What is the baseline value for the metric?	
Current benchmark	
Benchmark Algorithm version (with date)	
Benchmark algorithm comment	
Industry standard benchmark?	
Explainability of the Model (Does the new model improve the capability make decisions)	
Can inputs be traced back to outputs?	

4.5.6 Compute Resources

How is success measured and evaluated over time and what is the latest status.

Compute Resources	
What compute resources are expected to be used?	
CPU or GPU?	
Cloud or on Prem Resources?	
Cloud service provider?	
Why is cloud used or not used?	
Expected Cost of compute resources?	
Integrated with business systems?	

4.5.7 Human Resource Input

Who will be working on this use case development

Human Resources	
Exec / Manager Sponsor	
Product Manager	
Developer(s)	
Business Subject Matter Expert (engineering)	
External resources for development and other expertise	

4.5.8 External Vendor Collaboration

How will software vendors and developers collaborate or develop this use case?

External Vendor Collaboration	
What external vendors will collaborate on the project	
What role will they play	
Where have they demonstrated previous capability associated with, he uses case	
How will they access and securely hold data ***What is the expected outcome***	
For Open Source tools – licensing arrangements	

4.5.9 AI ML Model and Algorithm Development Process

In addition to the MLUCM, a basic machine learning development cycle was also developed as part of the project and shown at a high level in Figure 5.The use case prototype for alarm management that was developed will follow this life cycle through the early stages of developments, documented in Section 6

4.6 Summary

With a defined methodology and life cycle for machine learning use cases, as described in this section, the development of ML use cases for operations and control room applications will become more sustainable, risk-assessed and manageable. The specific use case is selected in Section 5 and the MLUCM is tested for this use case in Section 6.

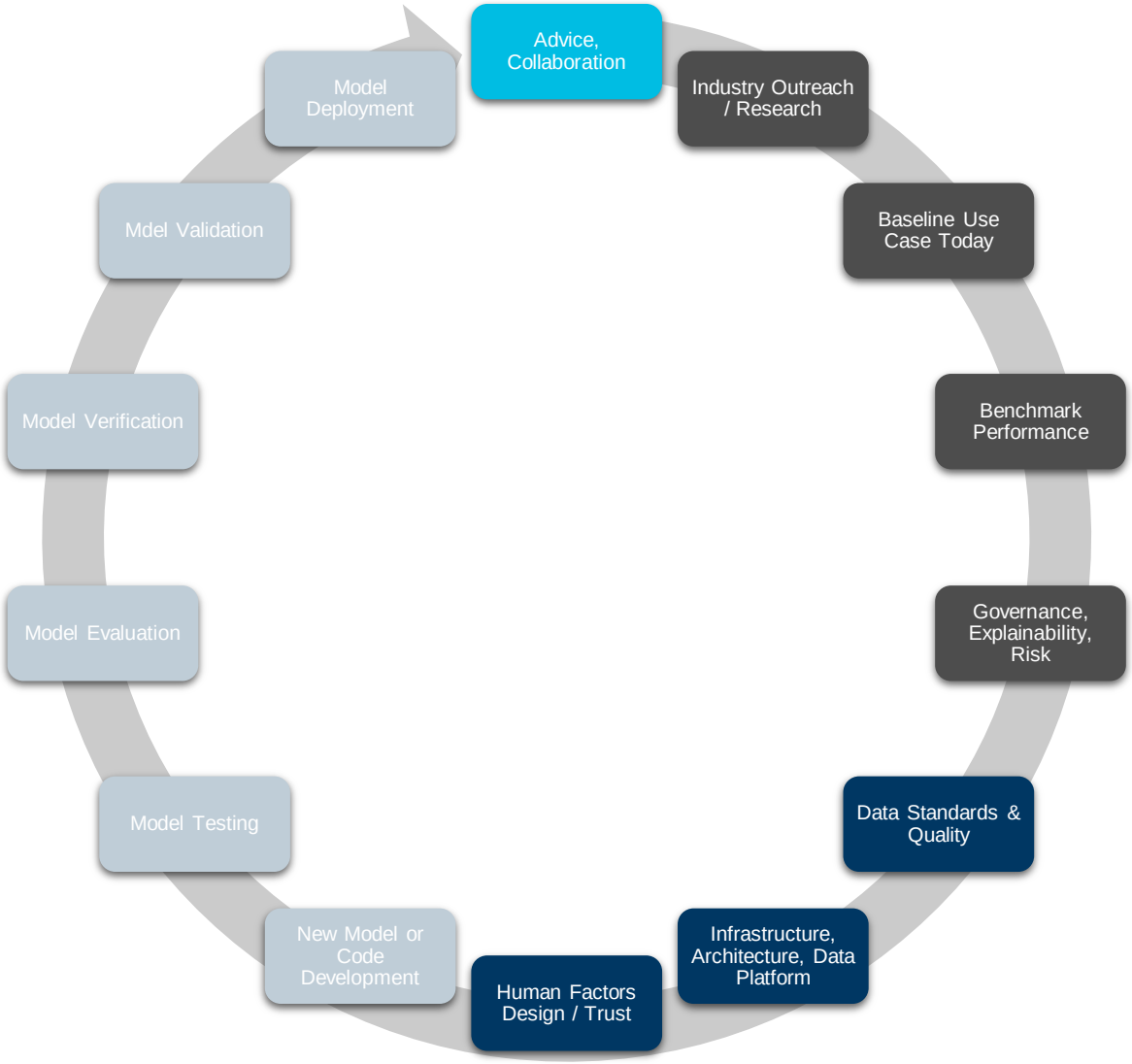


Figure 5 AI ML Development cycle for AI/ML use case development

5. Use Case Selection

5.1 Use Case Long List

A long list of potential use cases for the application of AIML was developed for assessment by the project team. The long list of use cases in Table 6 in Appendix A. The aim was to cast the net wide, to get all possible potential use cases into one list and to filter them to a short list of the most appropriate use cases for operations and the control centre domain. based on several criteria.

5.2 Use Case Short List

The long list of use cases (In Appendix A Table 6) were evaluated based on several criteria:

- Use cases with time series data were favoured given the range of data available and suitability to ML methods.
- Use cases with a large amount of time series data available were favoured, such as archive operational data.
- Use cases of high value and relevance to AEMO were preferred as AEMO would be project partners and data suppliers,
- Use cases that could be directly useful to other TNSPs in Australia were favoured, so that it could be potentially utilised by other network operators.
- Use cases that could provide the foundation for other use cases were favoured.

Four use cases were selected for potential advancement for development by the project team, to test the methodology and to attempt prototype development, shown in Table 4 - in order of preference:

Table 4 High Priority use cases selected for advancement for development through the methodology.

Priority Use Case	Use Case
1	Alarm Management / Event analysis
2	Constraint Process Automation
3	System Strength Analysis
4	Market price modeling and bidding behaviour

5.3 Alarm Management / Event Analysis

Goal	Reduce alarm noise levels, based on identified operator issues with alarms. Develop methods to identify alarm anomalies or system abnormities from alarm spikes and alarm chattering behaviour. Intelligent alarming – condensing multiple alarms from an event into one, log data + SCADA data. Utilise time series SCADA/EMS data for training. Potentially synchronise with other data sources. Use as a foundation for incident identification method and constraints automation method.
Baseline	Current rates of alarms per hour / day Number of spikes of alarms Chattering / noisy alarms within spikes.
Data of Input	Ability to use synthetic alarms dataset for testing and model development. Real network SCADA alarm data archive for model training and testing. Some domain expert operator knowledge to answer questions as required

Expected Outcome	Reduce overall alarm rates by identifying and hiding noisy low-priority alarms. Identify abnormalities, on the system based on SCADA alarm data, identify system incidents and link to past events.

5.4 Constraint Process Automation

Goal	Improve the management of constraints by identifying events and suggesting the market constraints to be invoked based on past invocations. Follow on from alarm management use case for incident identification. Constraint automation – which constraints invoked in past for events, potential to develop constraint equations automatically.
Baseline	Time taken / manual work involved in identifying and applying constraints during an event. Time taken to identify what has occurred. Errors made in application of constraints (if any)
Data of Input	SCADA alarm data archive, constraints log, market report information, PMU data, state estimator results. Some expert operator knowledge to answer questions as required. Initial focus on interconnector trip events
Expected Outcome	Automated, validated process for major market event identification and suggested constraints to be invoked by operator for the event.

5.5 System Strength Analysis

Goal	Identify system strength levels and deficiencies on the network based on operational scenarios, operational data, past events and calculation methodologies. Identify system strength issues ahead of time using forecast information and past operational scenario events. Linking events with system strength deficiencies to operational forecast.
Baseline	Short circuit levels on the network. System strength levels. System strength on the network based on existing calculation methodology and time spent making the calculations.
Data of Input	Network topology, generator status (state estimator solution), short circuit strength. Forecast information and pre dispatch information Some expert operator knowledge to answer questions as required
Expected Outcome	Automated process for system strength calculation in real time and predicted system strength issues based on operational forecast.

5.6 Market Price Modeling and Bidding Behaviour

Goal	Forecast and predict market prices ahead of time based on available information Early stage bidding behaviour modelling to understand price movements based on forecasts and operational conditions.
Baseline	Forecasts for market prices using existing methodologies and tools at AEMO. Actual market prices for past events.
Data of Input	Market data, forecast data, pre dispatch data. Some expert operator knowledge to answer questions as required
Expected Outcome	Forecasted ranges for prices ahead of time to allow engineers and operators to plan. Early-stage bidding behaviour models.

5.7 Selected Use Case

Given the time constraints for the project, the highest priority use case was selected for advancement and testing through the methodology. **This was the Alarm Management / Event Analysis Use Case.**

The alarm management/ event analysis use case was documented using the MLUCM (as outlined in Section 4). The results are shown in Appendix B. Having been assessed in the MLUCM – the use case was developed by the project team.

6. Use Case Development - Alarm Management / Event Analysis

6.1 Use Case Objective and Aims

The objective of this project task is to deploy machine learning techniques on time series alarm operational alarm data.
The aims of the use case prototype development as part of this project were:

- To give a forecast “look-ahead” of the likely alarm load based on current and recent past data.
- To automatically identify anomalies in normal alarm loads such as alarm “spikes”
- To identify the categories of alarms within the “spikes” such as chattering behaviour
- To find patterns within the alarm “spikes” that may be related and linked to past events.

Ultimately the algorithm or model develop should be capable of being able to distinguish between noisy alarm data (such as chattering alarms) and key event data (with breaker and generator operations).

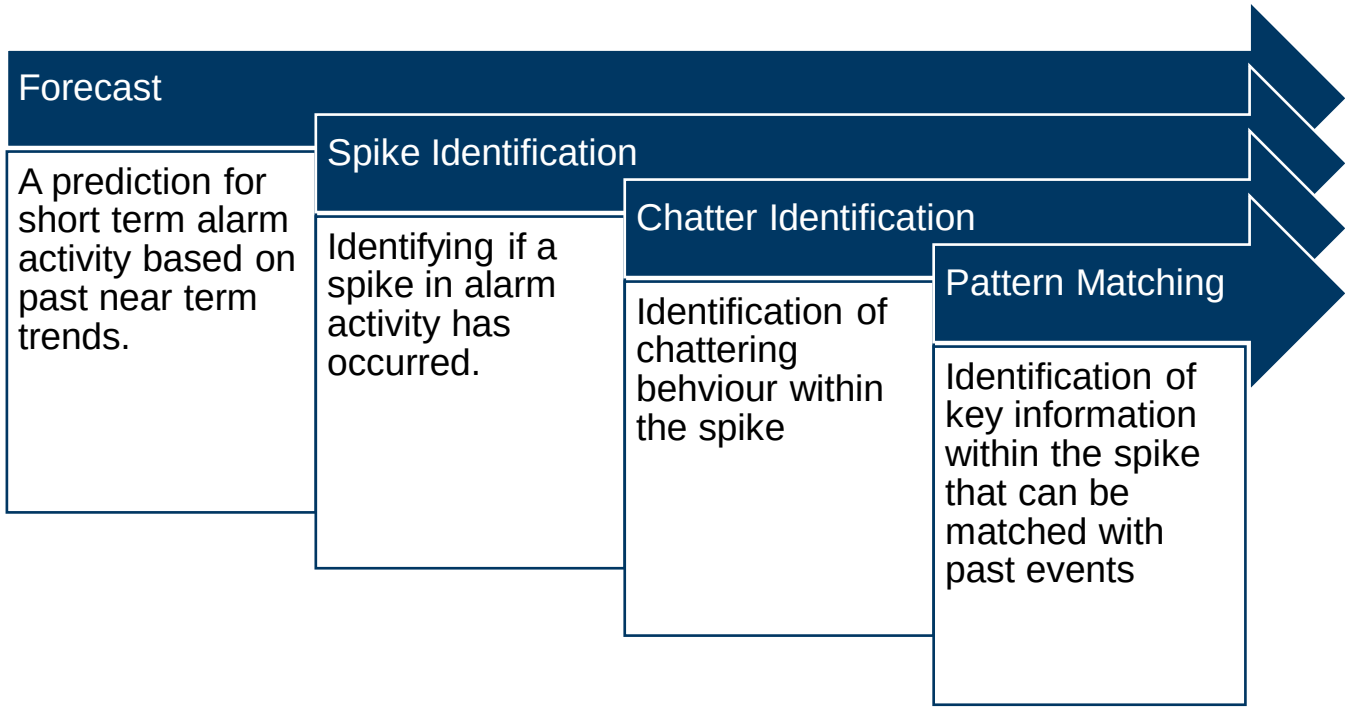


Figure 6 The four aims of the alarm management / event analysis AIML use case

6.2 Alarm Loads and Spikes

Operators in control rooms are tasked with analysing vast quantities of alarms that appear on their screens via SCADA. The majority of these alarms are basis status or situational information, that does not require intervention. However, a small percentage of alarms will require attention or may be a pre-cursor to an incident occurring. Well-designed alarm systems have rationalised alarm sets to filter out spurious information, to only display the most relevant information to operators. However, this is difficult to achieve

in practice and is exacerbated by the volume of new alarms from new technologies and assets that are being brought to the control centre operators' desks.

Generally, operators may have to assess hundreds of alarms per minute, with actions being required on a small percentage of those. When a major power system disturbance occurs, the alarm "load" will increase dramatically and the operator may be overwhelmed as they try to assess, diagnose, and mitigate the issue. See Figure 7 for an illustration of this effect.

The spikes in alarm load are clear indications of anomalous behaviour, however these spikes could be due to an actual event or disturbance on the system or due to a sensor or software system malfunction. An algorithm that automatically analyses and summarizes the key information would be very advantageous for decision making in real-time operations, allowing operators to separate signal from noise.

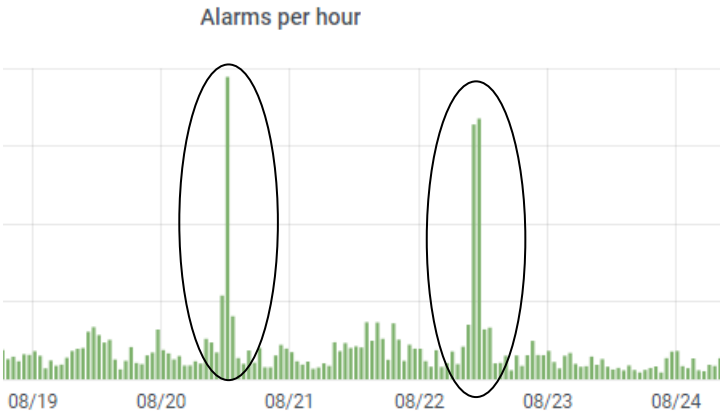


Figure 7 Example Alarm load trend showing baseline alarms and spikes

6.3 Datasets for Analysis

To develop the use case, a large amount of time-series alarm data is required to train the machine learning algorithm for the various aims, as each ML model will work to achieve a different output. Due to data security and sensitivity, the project team worked to develop the models with a synthetic alarm dataset and a real alarm dataset from AEMO, both datasets had a similar alarm schema.
This allowed flexibility in development while maintaining data security.

The synthetic dataset was made up of 5.78 million synthetic alarms representing one month of hypothetical timespan from an open source of information collated by Tommy Morris for his Industrial Control System Cyber Attach Datasets[7]

The AEMO test dataset was limited to one month of data for the purposes of protype testing, to be an equivalent size to the synthetic dataset.

6.4 Alarm Forecasting

The first method attempted for the use case was a basic forecast of alarm load. The aim here is an early warning system to notify operators with a distinct alarm if there is a high likelihood of a detected uptick in alarm activity.

[7] Source of Open-Source Dataset: Industrial Control System Cyber Attack Datasets: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

6.4.1 AIML Method Used

For alarm forecasting the aim was to take the alarms per minute and give a prediction of what the alarms per minute would be in a 1-to-5-minute period. This should allow an operator sufficient time to be alerted to a slowly developing disturbance. While this method will give a good general indication of alarm activity and its likelihood to increase or decrease, it will not predict contingent random events that occur. The long-term aim may be to identify anomalies in the alarm data as well as alarm forecasts, by linking with different datasets such as analog data, phasor measurement unit data.

The machine learning method used was Long-short term memory or LSTM. LSTM models are advantageous for time series and predictions as they do not rely on single inputs but can take in multiple datapoints or sequences of data as its input to predict the output.

6.4.2 Results on Synthetic Dataset

The LSTM model was tested on the synthetic alarm dataset (5.8 million alarms) and showed promising results. The performance of the algorithm was measured used the Mean Absolute Error (MAE) and the root mean square error (RMSE). For a dataset window of 16 hours (3 hours of testing data shown below in Figure 8, with input window of 15 minutes before and 1- and 5-minutes prediction, the RMSE was stable at between 28.4 and 29.3 and the MAE was stable between 21.95 and 23.02. Looking at the forecast versus actual in Figure 8 below gives a good indication of the efficacy of the algorithm, with the red line showing good predictive behaviour. It should be noted that these predictions are entirely independent of the underlying network model or the actual text in the alarm.

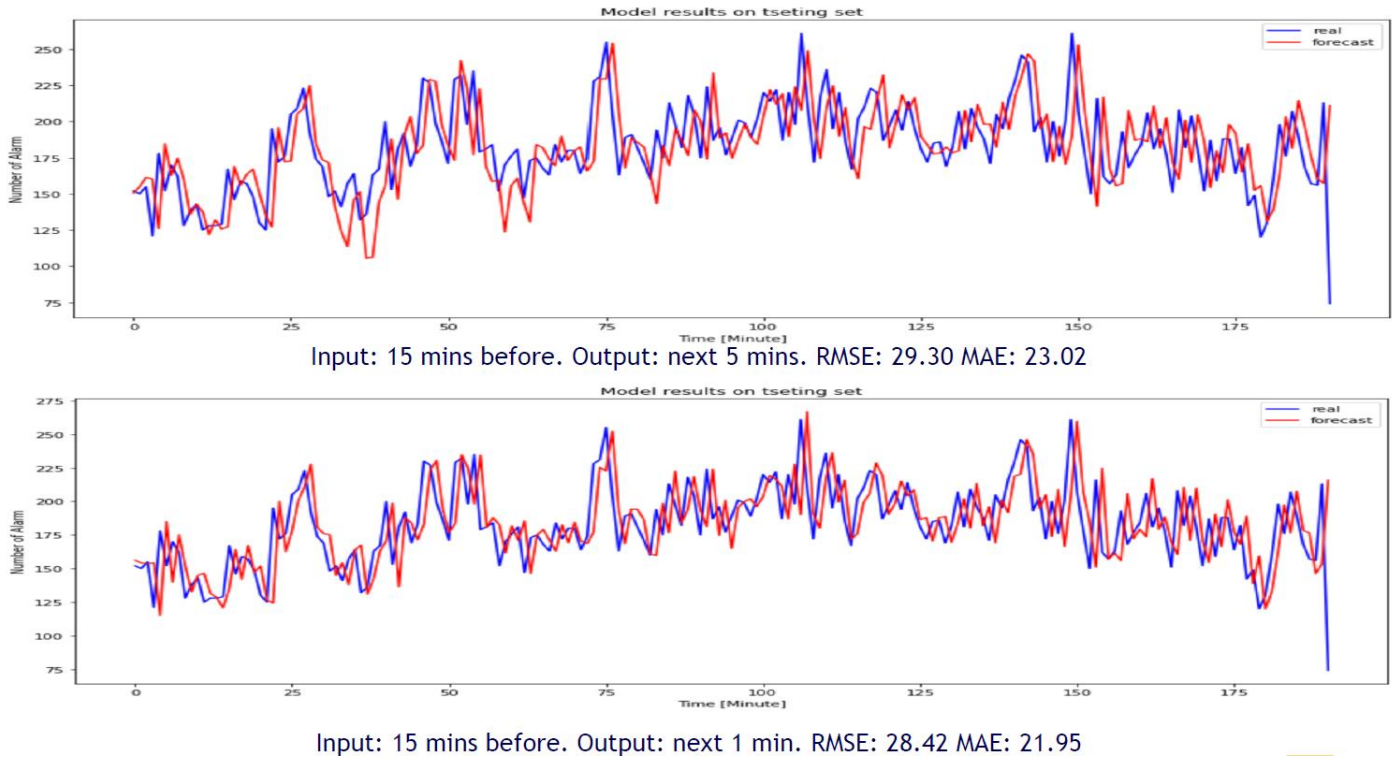


Figure 8 Results of the alarm forecast LSTM model on the synthetic dataset, showing good predictive performance on the synthetic dataset

6.4.3 Results on Real Dataset

The same LSTM alarm forecast model was tested on the real AEMO dataset for one day to test model efficacy. A varying set of input parameters were tested. The most promising results were achieved for input window of 20 minutes and output window of 1 minute, trained and tested on one day of alarm data. While the RMSE of 87 and MAE of 66 are higher than the synthetic dataset, analysing the prediction with the forecast visually shows promising results, although some refinement is needed to improve the forecast accuracy for alarm magnitudes and confidence range. The forecast versus real alarm load is shown in Figure 9.

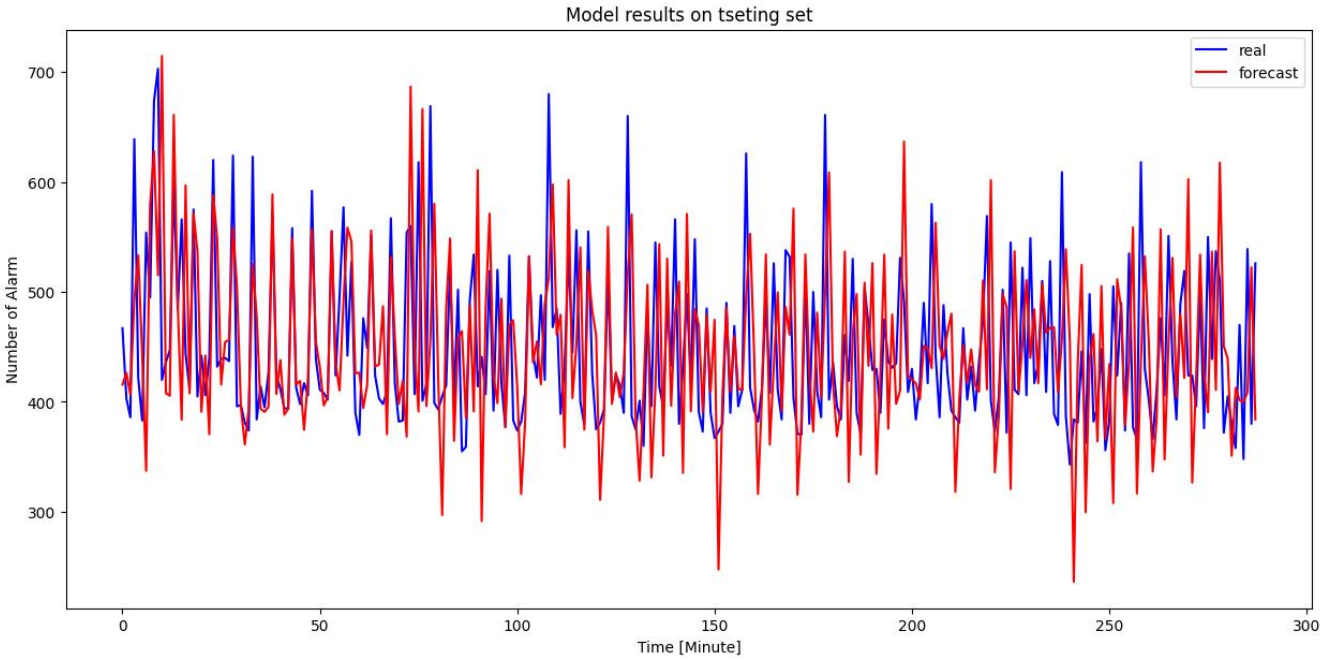


Figure 9 Results of the alarm forecast LSTM model on the real AEMO dataset for one day (testing set only shown), showing good predictive performance. Input window of 20 minutes, output prediction of 1 minute.

The prediction error rates tended to increase as the training testing dataset was increased, showing that there may not be advantages in using long term alarm data for short term predictions, however more testing and tuning is required for this model. The model was applied to two weeks of data as a test and is shown in Figure 10 with low visualisation resolution.

In future iterations and developments of the algorithm, the project team will examine accuracy for positive and negative spikes, and how to minimize incorrect forecasts and to minimize the impact of false positive predictions.

6.4.4 Baseline Performance:

Input 20 mins, Output 1 minute, RMSE: 87, MAE 66.5

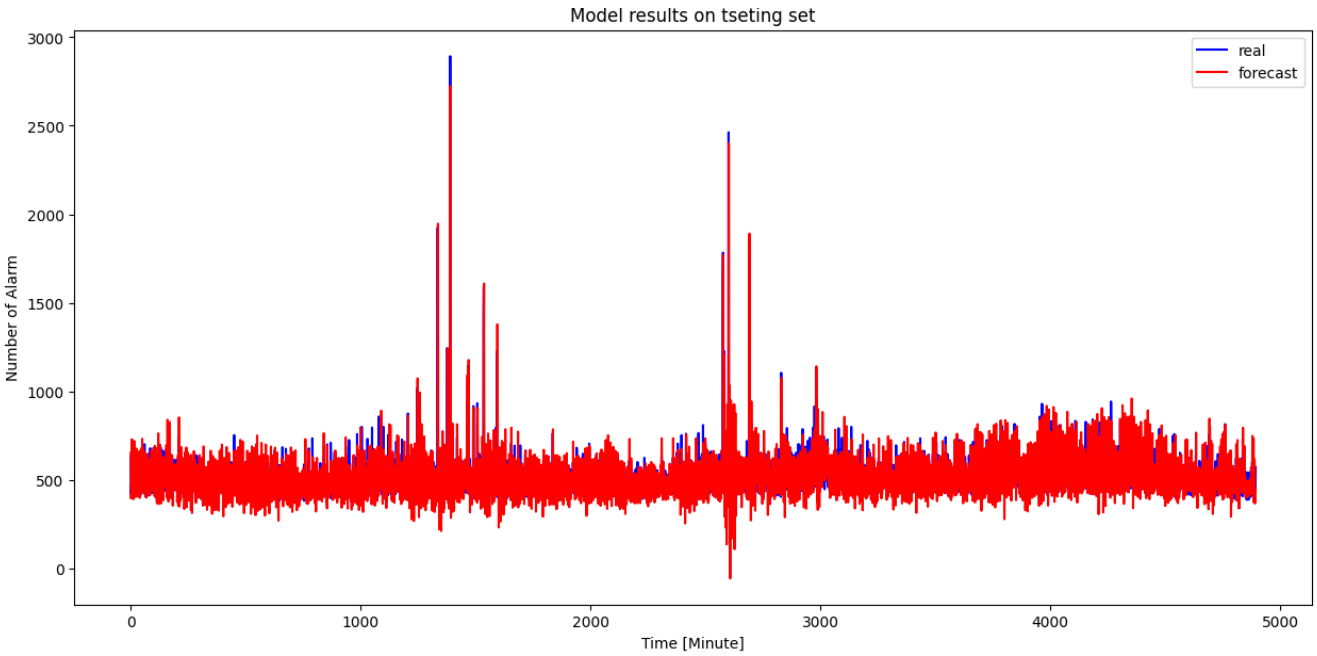


Figure 10 Results of the alarm forecast LSTM model on the real AEMO dataset for two weeks of alarms (testing set only shown), showing predictive performance. Input window of 20 minutes, output prediction of 1 minute.

6.5 Alarm Spike Detection

The second method aimed to identify abnormal activity in alarm loading. The most efficient way to identify anomalies in alarm loads is to identify what are called “alarm spikes”. These are deviations in alarm activity above the normal background alarm rates. An effective alarm system would automatically identify alarm spike activity when it occurs, so that it can be further analysed.

6.5.1 Baseline Statistical Analysis

To assess the statistical baseline, it is necessary to group the alarms by minute (more or less granular groupings can be applied, but 1-minute bins are most effective). The baseline rate is simply a calculation of the number of alarms per minute in a rolling 30-minute window. A rolling window is used, as alarm load can vary depending on the hour of day or day of week. Alarms per minute that are above a certain standard deviation threshold can be considered “spikes”. By running a mathematical statistical analysis to identify spikes based on standard deviation, the spikes can be labelled as a baseline.

6.5.2 Baseline Statistical Analysis on Synthetic Dataset

In Figure 11 the spikes are shown for the synthetic dataset for one month of data. It can clearly be seen that spikes with breakouts of standard deviation of 10+ will cause data overload that may lead to errors, panic, and mistakes by operators, or critical information getting lost. This 10 std dev number is based on experience and high-level analysis of the data. It will not necessarily apply for all alarm datasets, which will have their own rhythm and contexts. Additionally, some other basic statistical analysis and reporting was easily carried out on alarm data as shown in Figure 12

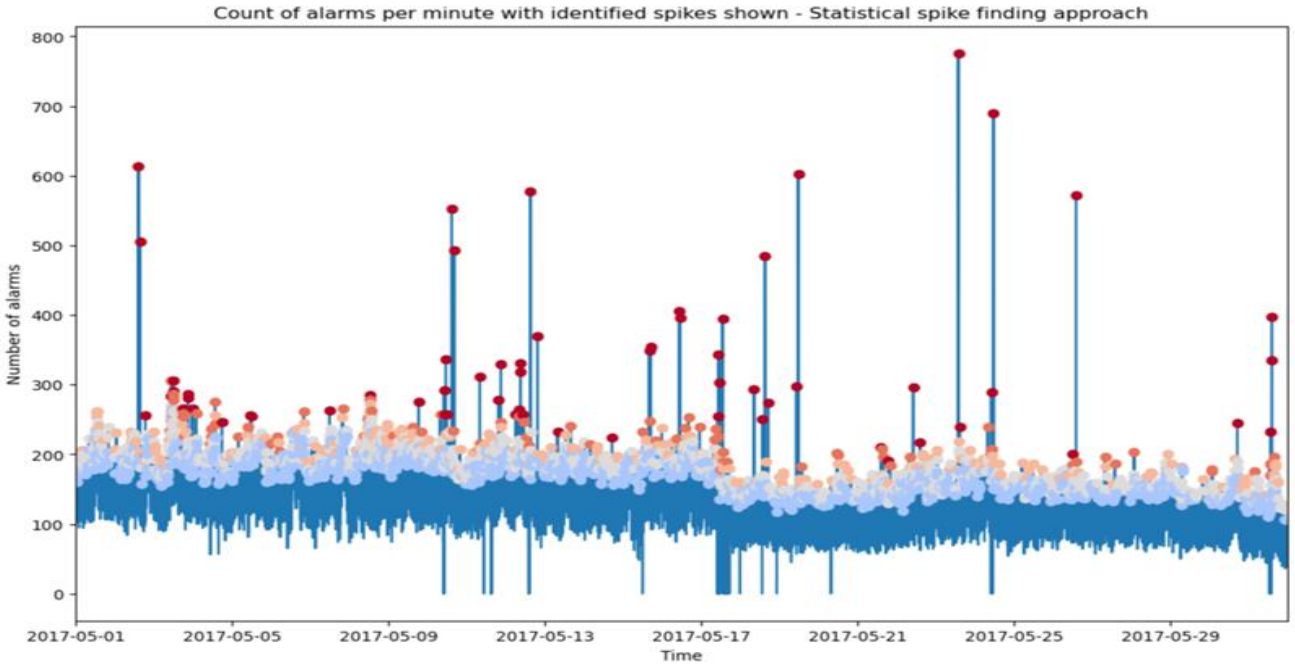


Figure 11 Statistical analysis of alarm load and spikes on one month of synthetic data.

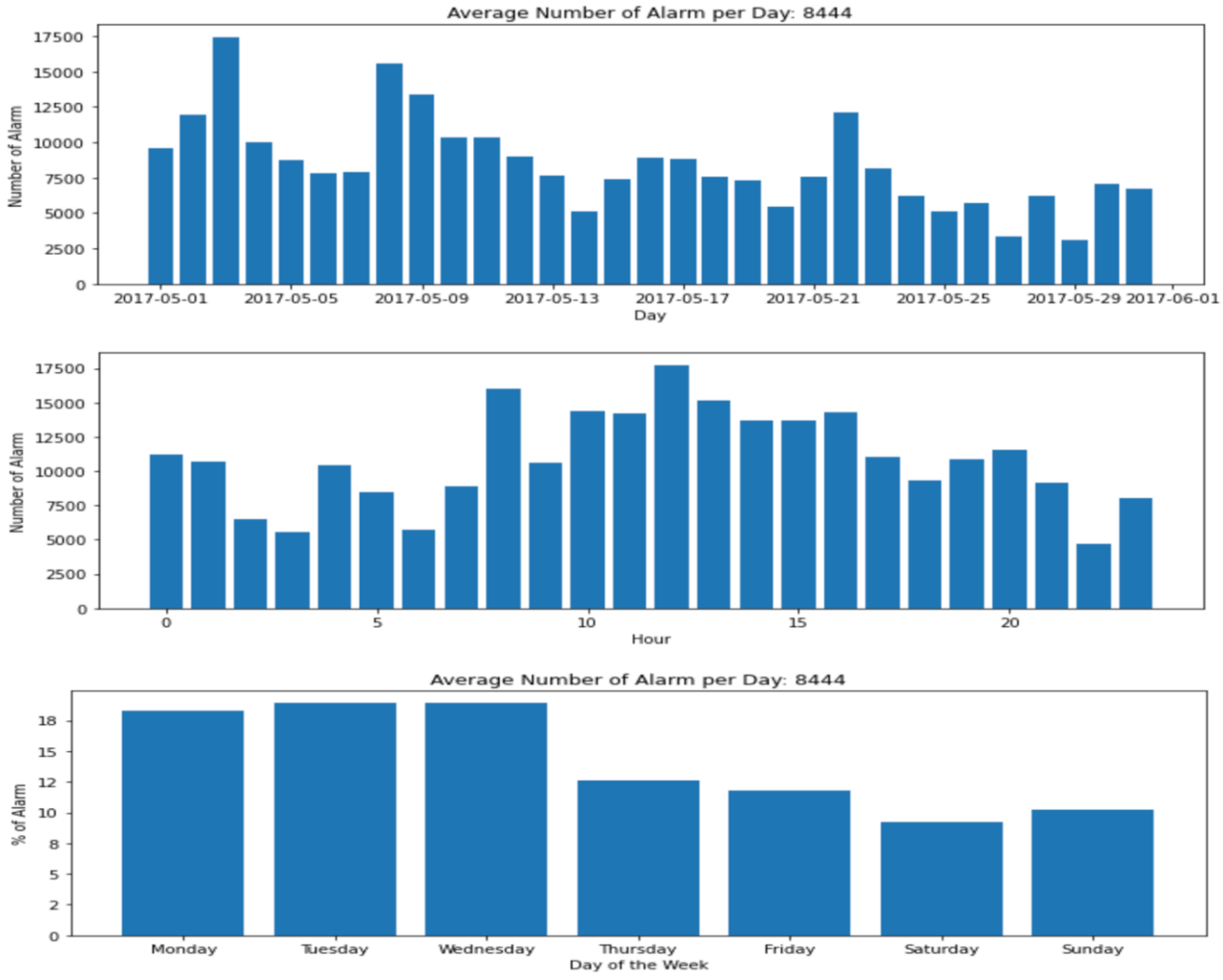


Figure 12 Top: Alarms per day, middle alarms per hour of day, bottom: Alarms per day of week

6.5.3 Baseline Statistical Analysis Real Dataset

The same model for statistical analysis was run on a real data set with one month of alarms. The result is shown in Figure 13. There is clear and obvious spiking behaviour above 15 standard deviations which indicate abnormal alarm activity. This spiking behaviour of up to 3500 alarms per minute may cause confusion, data overload, for the operators, which may result in a loss of situational awareness, unless they are suppressed, silenced or filtered out.

Figure 14 shows the number of alarms per minute in spikes relative to the number of standard deviations. Standard deviation of 3+ in the dataset indicates spiking activity.

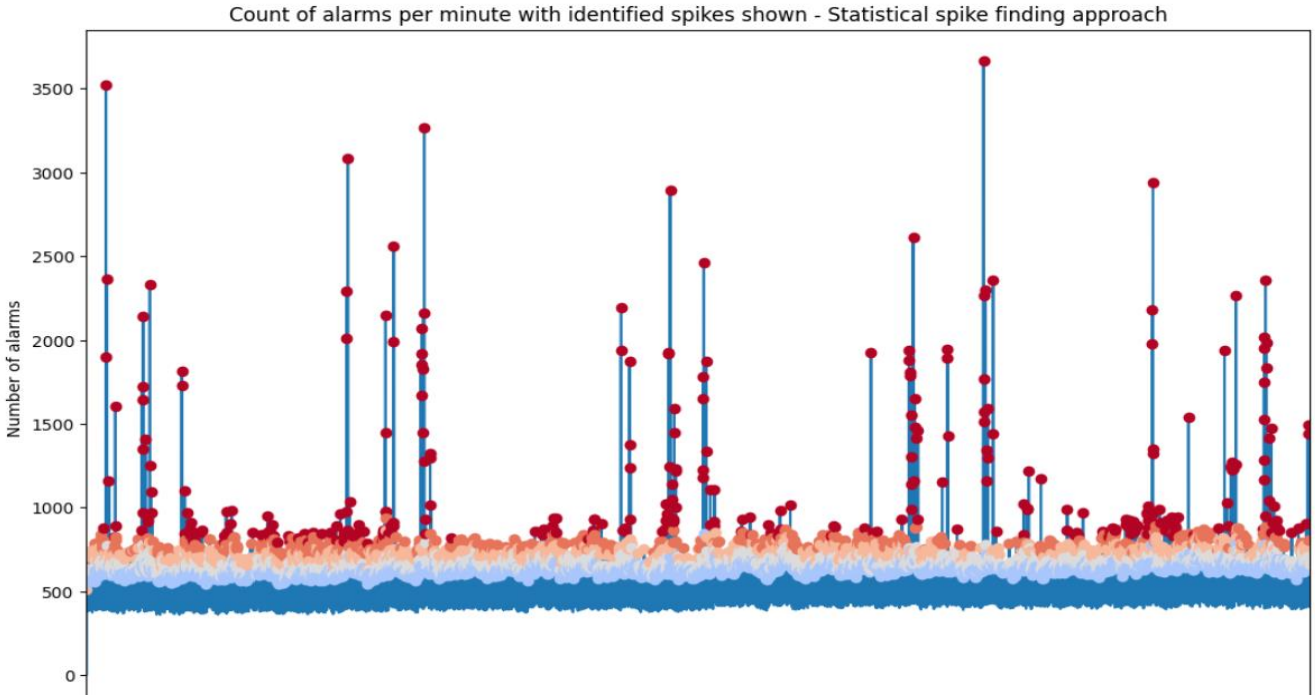


Figure 13 Baseline statistical analysis of the real alarm dataset for one month of alarm data.

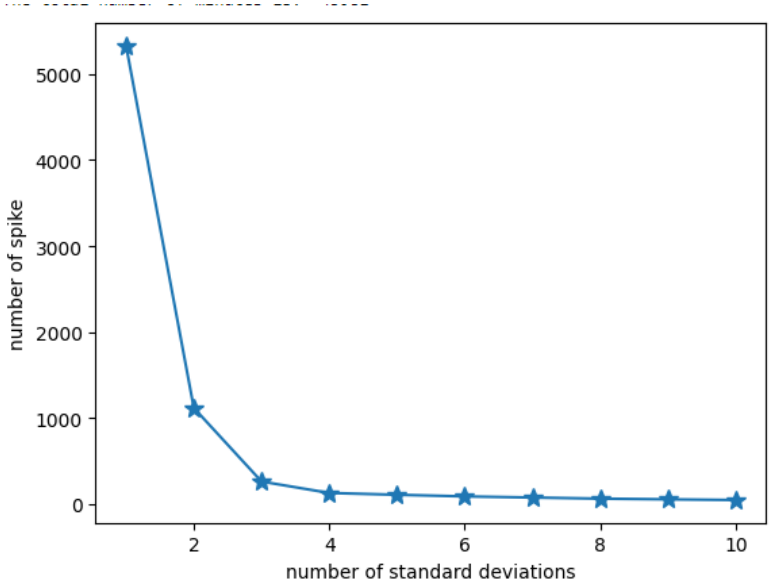


Figure 14 Number of alarms per minute spikes relative to standard deviations. 3+ Standard deviations are a good proxy for spiking behaviour

6.5.4 AIML Method Used for Spike Detection Algorithm

Using rolling statistical analysis to identify spikes may not be computationally efficient and so an ML approach was tested that would be trained on alarm data with labelled spikes (from the statistical analysis). A classical neural network was used for this algorithm with varying weights for the inputs to the neural network.

6.5.5 Results on Real Dataset

The results of the application of the spike detection machine learning model were promising for a baseline model. The results are summarised in Table 5. The key point to note is that accuracy of the model of identifying instances of alarms per minute that are not spikes is very high. The model accuracy of accurately identifying spikes is also good, however, there are for a far lower number of instances of spikes at higher standard deviations, so the sample size is relatively small.

The computation time for the model is also very low. The model can give results in less than 1.5 seconds once it is trained. Some further work is required to tune the model and train it on a wider array of data.

6.5.6 Baseline Performance

Table 5 Results of the spike detection baseline model

Spike Standard Deviation	Spike Weight	Number of Actual Spikes in Test Dataset	Correct Negative Prediction	Correct Positive Prediction	Incorrect Negative Prediction	Incorrect Positive Prediction	% Of Correct Spike Predictions	% Of Correct Non Spike Predictions	Time for Test (seconds)
3+	2	285	8505	214	71	1	75%	96.75%	1.3
6+	1	25	8766	23	0	2	92%	99.72%	1.5
11+	1	10	8781	8	2	0	80%	99.89%	1.4
16+	1	2	8789	1	1	0	50%	99.98%	1.5

The confusion matrix is shown below in Figure 15 this is a visualisation method for machine learning showing the results of a test of an ML model for binary predictions. The top left quadrant indicates that the datapoint was negative (no spike in alarm per minute) and the prediction was negative (desired). The top right indicates a negative datapoint and positive prediction (this is undesired behaviour). The bottom left quadrant indicates a positive datapoint (spike) and a negative prediction (undesired) and the bottom right indicates positive datapoint (spike) and positive prediction (desired).

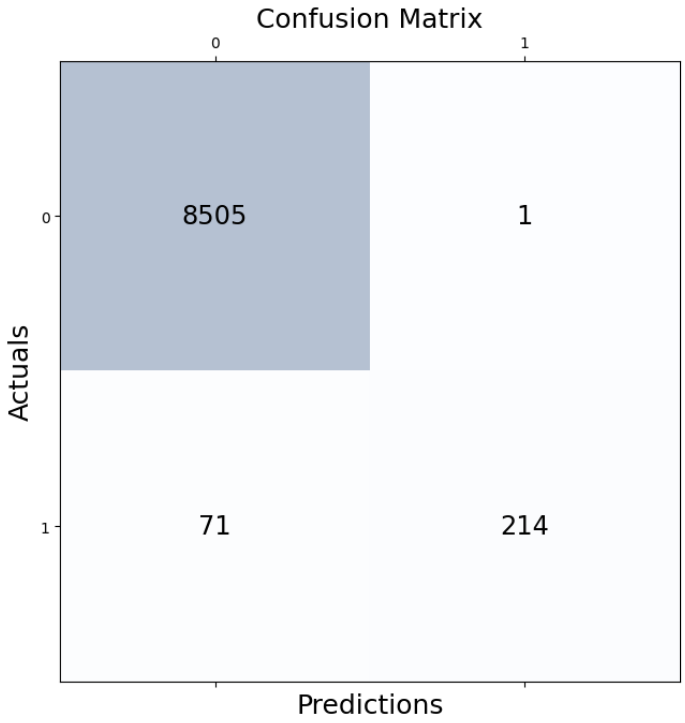


Figure 15 The confusion matrix which classifies graphically the results of the model for spike detection of std dev 3+ with a weight of 2. Correct positive predictions are 214 of 285 total positives.

6.6 Alarm Chatter Detection

A common feature of alarm spikes is the so-called chattering behaviour. This is characterised by the same alarm appearing multiple times in rapid succession due to flickering status changes, e.g., ON-OFF-ON-OFF etc. Chattering alarms can increase the number of alarms to hundreds or thousands per minute and cause data overload, confusion, and can mask real events and data points. Chattering behaviour is very undesired in real time operations and requires the operator to manually disable, filter or supress the alarm. The alarm must be reported to an external entity to be corrected. Chattering, repetitive alarms offer little or no value to situational awareness and can actively disrupt effective operations.

The standard rule to identify chattering behaviour is if an alarm appears more than 6 times in one minute it can be deemed to be chattering. That is if the same alarm appears 3 times with changes of status.

Chattering behaviour can appear in alarms loads that are not spikes, but they can be more easily managed by the operators as part of normal activities.

6.6.1 Alarm Chatter Labelling Real Alarm Dataset

Alarm chattering labelling is simply the application of the deterministic rules for chattering as described above. The rules were programmed into an automated script and applied to the dataset, this is an efficient method of labelling for chattering, and the rules can be adjusted to fit the context (for example in some systems less than 3 per minute alarms may constitute chattering behaviour The rule-based labelling on a normal computer can be slow. On the real alarm dataset for one month of data, 77% of the alarms were chattering and it took 45 minutes to label the dataset.

6.6.2 AIML Method Used

A back propagation neural network (BPNN) was used for chattering detection, with a variety of input layers depending on the alarm data fields. A BPNN is defined as a model that passes back the error detected in

the output to the input, so that the weights can be adjusted to achieve the objective. It acts as a feedback loop for the training, similar to an optimization algorithm

An LSTM model was also tested but proved to be less accurate when compared to the BPNN method.

6.6.3 Results on Real Dataset

The BPNN model was trained and tested on one month of real alarm dataset. The results were very promising. For one month of data the testing accuracy was 89%, meaning in 89% of cases the chattering behaviour was correctly identified by the ML model when compared to the rule-based labelling for chattering.

6.6.4 Baseline Performance

1 month alarm dataset 22 million alarms.
Chatter identification accuracy: 89 %
Testing time 145 seconds

6.7 Pattern Matching - Spike Analysis

When the alarm spikes have been identified and chattering alarms filtered out, it will be necessary to intelligently assess the alarms in the spikes – to aid operator decision support and guide them quickly to the source of the system disturbance that generated the spike. The most important step in alarm spike analysis is to automatically identify what is happening within the spike and condense the language in an understandable way.

6.7.1 AIML Method Used

The alarms as presented are usually not labelled or contain additional information in their basic form, and so there is little contextual information beyond the raw alarm text. To get over this limitation, an unsupervised clustering algorithm is required that could use machine learning. The approach taken was to combine the text in the alarms of each spike separately and to analyse the text in the entire alarm spike to gain insights. The approach used consisted of:

- Combining the text from the spikes into a single “document.”
- Using the Doc2Vec python package which uses neural networks to convert the text in the documents into a sparse vector representation. There should be one vector for each alarm spike.
- Use a “self-organising map” (SOM) which is neural network that clusters the spike vectors visually, which should show patterns of closely aligned words within the spikes.

6.7.2 Results on the Synthetic Dataset

The clustering / self-organising map method was applied to one month of synthetic alarm data, using different inputs from the alarm field.

The first test applied the AOR filed which is the Area of Responsibility. The algorithm identified the top 3 AOR combinations in each alarm spike and plotted them on the self-organising map based on similarity.

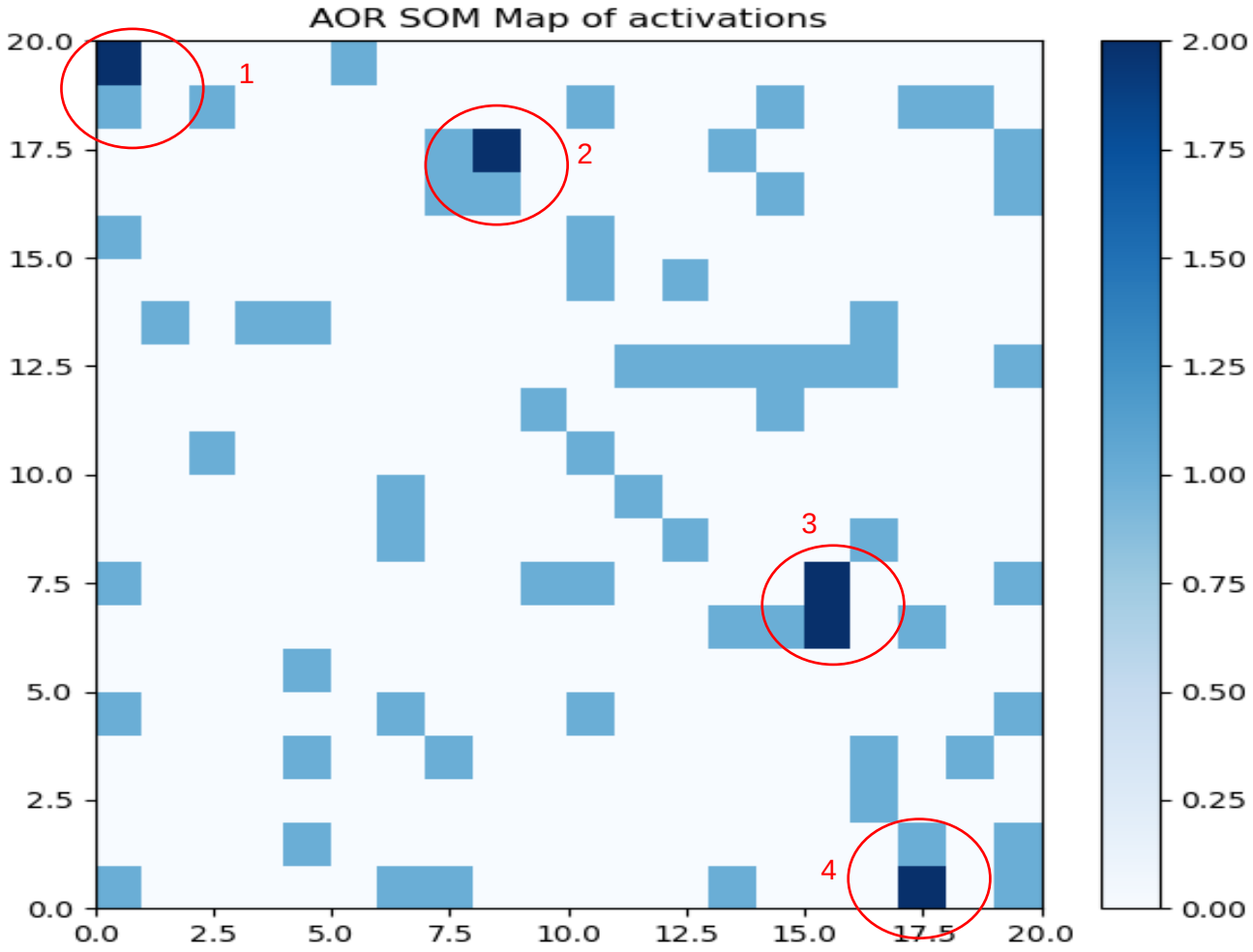


Figure 16 Example of an SOM for the AOR alarm field showing intensity of the patterns in darker shade of blue.

This is shown in Figure 16, with red circles indicating high intensity of combinations which require further examination. This is shown clearer in Figure 17, where the dark blue pattern indicated in circle 1 represents a clustered combination of the AOR: [TOCBR, TOCBM,TOCGR].
In operator terms this would mean that an event involving these three areas would cluster in this position.

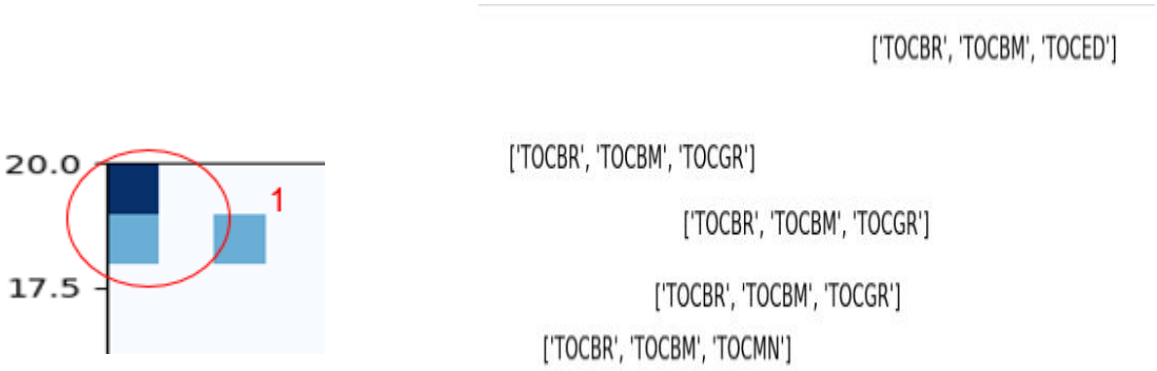


Figure 17 The corresponding text combinations for circle 1 showing a clustering of three AORs [TOCBR, TOCBM,TOCGR]

The same clustering principle can be applied to combinations of substations, device names, and other alarm text words. From example combining AOR, Substation, DeviceType and Device in a vector and clustering on an SOM gives the result in Figure 18.

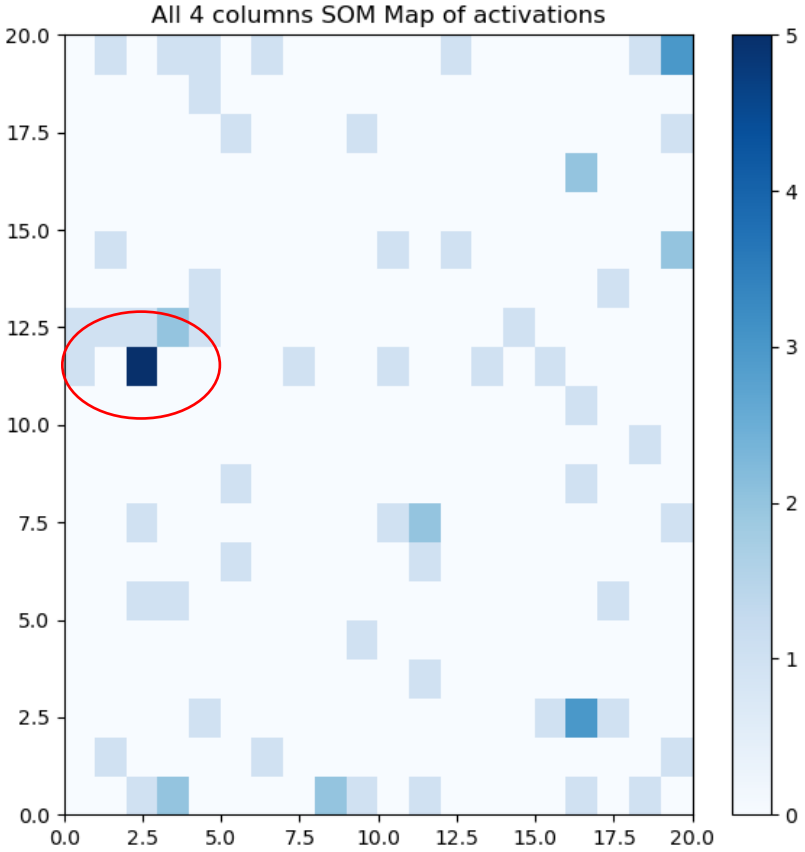


Figure 18 SOM of alarm spikes in the synthetic dataset for the four alarm attributes AOR, Substation, DeviceType and Device. The red circle indicates an intense cluster of activity.

Looking into the red circle of intense activity in Figure 19 shows a dense concentration of the Device Relay, in AOR TOCBM and TOCGR. This can add some intelligence to decision making if developed and applied correctly.

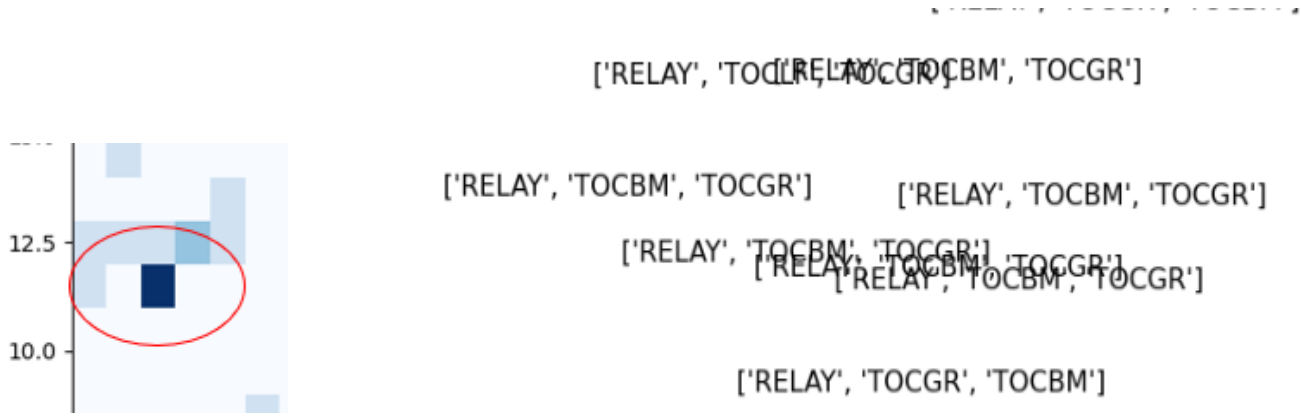


Figure 19 Most intense area of the SOM cluster for the four alarm attributes

6.7.3 Results on Real Dataset

This model was not applied on the real alarm dataset due to the early maturity in the development and the need for further exploration and tuning to on the synthetic dataset to be appropriate for use on the real alarm dataset.

Further research will be carried out in future phases of the project and the model will be applied and reported on at a future date.

7. Conclusions

Overall, the project was successful by introducing a methodology for the assessment of AI. ML use cases in the system operations and control room context and for the development of prototype, low technology readiness solutions that are being run successfully on AEMO systems.
Some other notable conclusions from the research on the project are detailed below:

- Given the large quantity and availability of (primarily time series) data in control room and system operations functions generally - machine learning is a potentially powerful application if the correct use cases are defined.
- System operators can be overwhelmed by the large quantities of data beyond the human capability to cognitively filter and supress information in real time.
- Traditionally, utilities do not have the capability to exploit the vast quantities of data available to them, due to lack of expertise on machine learning models, development overheads, human resources and compute power resources.
- These barriers are being reduced in recent years, with the availability of advanced compute power and the development of data science and machine learning specialists and teams within utilities.
- However, beyond load and weather forecasting, which has been established for many decades, machine learning has had limited breakthroughs in control room and network and market operations functions.
- There is a renewed push for utilities across the world to leverage the expertise available in the wider community, vendor capability, increased compute power form on-premises and cloud resources and most importantly the quantities of data available to deploy machine learning to solve the challenges of network operation.
- There is currently no standardised methodology for developing machine learning and artificial intelligence use cases in the energy sector. (In addition, standardisation frameworks for use case development are not very prevalent through the industry.) This is a big challenge given the criticality of the network and the risks posed to real time operations of insecure or untrustworthy systems.
- A structured approach to developing machine learning use cases for control room operations is required.
- This project task aims to establish a first of its kind methodology for the development of machine learning use cases for system operations and control room operations.
- The methodology leverages existing materials in the wider industry which are relevant and important to the control room operations use cases such as:
    - EPRI and IEC Use Case Methodology
    - NIST AI Risk Framework
    - Machine Learning Life Cycles
- The methodology has nine core components that map to the nine key challenges that are faced with machine learning development.
- The objective of the methodology is for product managers and developers to fill in advance of development and to refer to throughout the machine learning project development life cycle.
- There is an extensive section on risk management (from the NIST AI risk framework), all of which may not be relevant, but much of which should be considered during project design.

- Baselining and benchmarking form a key part of the methodology, establishing metrics and values to be tracked during development and improved upon over time as innovation in the project develops. These benchmarks can be industry standards for ML innovations.
- To tests the methodology a use case for alarm management and event management was developed by the project team. The use case had four main elements consisted of
    - Alarm Forecasting
    - Alarm Spike Detection
    - Alarm Chatter Detection
    - Alarm Pattern Matching
- The elements of the use case were developed, trained, and tested using both synthetic alarm datasets and real alarm datasets contributed by AEMO.
- For alarm forecasting a long-short term memory (LSTM) model was developed which showed good predicative quality with an input of 20 minutes a forecast output 1 minute ahead. A baseline error rate was established that can be improved upon with further tuning.
- For alarm spike detection, a standard neural network model showed high degrees of accuracy in detecting so-called “alarm spikes” when given rule-based inputs of number of standard deviations above a baseline average alarm per minute rate. It had very high accuracy rate in detecting non-spike and reasonably good performance off smaller sample size for actual spikes.
- For the alarm chatter detection element, a Back Propagation Neural Network showed promising performance when tested on real data, by being able to detect 9 out of 10 chattering alarms when trained on a dataset that identified chattering behaviour through rules-based approach.
- In all cases the speed of performance of the machine learning models was very favourable when compared with standard statistical or rules-based approaches, however more work is required to deploy the models on faster servers to get a true reflection of the computational saving,
- The alarm pattern matching element was developed and tested on the synthetic dataset, it showed some good early promise on the synthetic dataset in visually clustering and grouping similar alarm text. It will be trialled on real datasets in future developments of the project.

8. Recommendations

The research aligns with the 2021 CSIRO G-PST research plan for topic 3 control room of the future. The data models and streaming pillar and the control room applications pillar. Given the long development life cycle of AIML projects, it was prudent to begin early-stage methodological development work and early-stage use case developments.

The stage 2 research project does not focus directly on model management or the digital twin, this will be developed by AEMO, CSIRO and EPRI further in other related work activities in 2023 and 2024.

The recommendation is to continue to develop this project topic area with further development projects in 2023-4 and beyond. The work in stage 2 was very low technology readiness level and it will only be improved with continued development, training, testing and validation before it can be deployed in real time control rooms.

9. Outstanding Research Activities

The next step in the use case development process, is to continue to develop the algorithms off the baselines established in this project. The project team are fully engaged in the issues with key collaborations with between EPRI, CSIRO, AEMO and RMIT. Beyond the alarm management and incident identification use case, the aim is to develop the next highest priority use cases associated with major incident identification market constraints management and improvements.

The benefit of using real data is that the network information can also be incorporated, and the real cause of the spikes can be correlated with reports and information in other databases, such as incident report logs.

If the algorithms of the four elements of this project can be tuned to be even more accurate using real data, they can be used to develop further automation within the operational environment, such as automated report generation or automated market constraint invocation and processing. They are also generalisable to be deployed in other operational and control room contexts.

Appendix A

A-1 Use Case Long List

The 60 use cases in the initial curated list are shown in Table 6, curated from lists of use cases from Cigre, EPRI and AEIC (Association of Edison Illuminating Companies) a collaborative body of electrical utilities in the USA. From this list - the four high priority use cases were selected by the project team for further development, see Section 5.

Table 6 Table of the long list of use cases for potential development with AIML applications

Source	Group	Use Case	Relevance to AEMO	Value to AEMO	Relevance to TNSPs	Time Series	Selected as High Priority
Cigre	Forecasting and Risk Assessment	Load forecasting (consumer behaviour)	3	2	1	Yes	
Cigre	Forecasting and Risk Assessment	RES (Renewable Energy Sources) forecasting	3	2	1	Yes	
Cigre	Forecasting and Risk Assessment	Outage Management and Planning	2	2	3	No	
Cigre	Forecasting and Risk Assessment	Frequency risk and uncertainty management	3	2	1	Yes	
Cigre	Forecasting and Risk Assessment	Impact of large weather events	3	2	3	No	
Cigre	Forecasting and Risk Assessment	EV (electric vehicle) and battery storage integration	3	1	1	No	
Cigre	Forecasting and Risk Assessment	Dynamic line rating	2	2	2	Yes	
Cigre	Grid Monitoring	Transmission grid state estimator	3	2	3	No	
Cigre	Grid Monitoring	Asset monitoring / failure prediction	1	2	3	Yes	
Cigre	Grid Monitoring	Alarm management	3	3	3	Yes	1
Cigre	Grid Monitoring	TSO-DSO coordination & cross-border coordination	3	3	3	No	
Cigre	Grid Monitoring	WAMS (wide area monitoring systems) / PMU (phasor measurement units) data analysis	3	2	1	Yes	
Cigre	Operations Processes	Congestion management	3	2	1	Yes	
Cigre	Operations Processes	Voltage management (steady state)	3	2	2	Yes	
Cigre	Operations Processes	Generation redispatch	3	2	1	Yes	

Source	Group	Use Case	Relevance to AEMO	Value to AEMO	Relevance to TNSPs	Time Series	Selected as High Priority
Cigre	Operations Processes	Dynamic security assessment (transient, voltage, frequency, small signal stability) System Strength	3	3	1	Yes	3
Cigre	Simulation	Data augmentation (operational scenario, synthetic data generation)	2	2	2	No	
Cigre	Simulation	Data annotation / labelling	2	2	2	No	
Cigre	Simulation	Real time digital simulators	3	2	2	Yes	
Cigre	Simulation	Operator training simulator	2	2	2	Yes	
Cigre	Market Management	Unit commitment	3	2	1	Yes	
Cigre	Market Management	Economic dispatch / redisptatch	3	2	1	Yes	
Cigre	Market Management	Reserves settings	3	2	1	Yes	
Cigre	Market Management	Consumer elasticity analysis	3	2	1	Yes	
Cigre	Market Management	Market price modelling and bidding behaviour	3	3	1	Yes	4
Cigre	Unplanned Extreme Emergency Events	Restoration / blackstart decision support	2	3	2	No	
Cigre	Unplanned Extreme Emergency Events	Event analysis / Constraint Automation	3	3	3	Yes	2
Cigre	Unplanned Extreme Emergency Events	Protection, special protection scheme co-ordination	1	2	3	No	
Cigre	Reporting Assistance	Knowledge extraction	3	3	3	No	
Cigre	Reporting Assistance	Report generation	2	3	2	No	
Cigre	Reporting Assistance	Voice-based assistance / natural language processing	2	2	2	No	
Cigre	Reporting Assistance	Digital assistant	2	2	1	No	
Cigre	Reporting Assistance	Augmented reality	1	2	1	No	
Cigre	Reporting Assistance	Visual detection and analysis	1	1	1	No	
Cigre	Reporting Assistance	Voice or video captioning	1	2	1	No	

Source	Group	Use Case	Relevance to AEMO	Value to AEMO	Relevance to TNSPs	Time Series	Selected as High Priority
Cigre	Reporting Assistance	Workforce management	1	1	3	No	
Cigre	Reporting Assistance	Process or workflow tracker (tracking actions against process with suggested clean-up steps)	3	2	1	No	
AEIC		Vegetation Management	1	2	3	No	
AEIC		Transformer Optimization using ML	2	1	2	Yes	
AEIC		Image Analytics Asset Inspection Strategy	1	1	3	No	
AEIC		Pad mount Transformer Vehicle Hit	1	1	3	Yes	
AEIC		Drone and Image Analytics	1	1	3	No	
AEIC		Non-Priority Reject Pole Prioritization Model	1	1	3	No	
AEIC		Power Transformer Asset Management	1	1	3	No	
AEIC		Transmission Fault Cause Detection	3	2	3	Yes	
AEIC		Fault Location Analysis	3	2	3	Yes	
AEIC		Distribution Linear State Estimation	1	1	1	No	
AEIC		Imminent Outage Prediction	2	2	2	Yes	
AEIC		Energy Theft / Revenue Protection	1	1	1	Yes	
AEIC		Outbound Notification Analytics	1	1	1	Yes	
AEIC		Small Damage Claim Outcome Prediction	1	1	1	No	
AEIC		Customer Analytics – Arrears Reduction	1	1	1	No	
AEIC		Metered but Unbilled	1	1	1	Yes	
AEIC		Advanced Metering Infrastructure (AMI) Anomaly Detection	3	1	1	Yes	
AEIC		Battery Energy Storage Charge Discharge Strategies	3	2	1	Yes	
AEIC		AI ML for Cyber Security	3	2	3	Yes	
EPRI		Grid-Interactive Smart Communities	2	1	2	No	
EPRI		Energy System Resiliency	2	1	2	No	

Source	Group	Use Case	Relevance to AEMO	Value to AEMO	Relevance to TNSPs	Time Series	Selected as High Priority
EPRI		Environmental Impacts	2	1	2	No	
EPRI		Intelligent & Autonomous Power Plants	1	1	1	Yes	

Appendix B

B-1 Alarm Management / Event Analysis Use Case Applied to the MLUCM

Name of Use Case

This section details basic background information

Use Case Identification		
ID	***Domain(s)***	***Name of Use Case***
1	***System Operations***	***Alarm Management / Event Analysis***

Version Management

This section details basic version control management information

Version Management					
Version Management Changes / Version	***Date***	***Name Author(s) or Committee***	***Domain Expert***	***Area of Expertise / Domain / Role***	***Approval Status draft, for comments, for voting, final***
0.1	***23rd June 2023***	***CSIRO/EPRI/RMIT/AEMO***	***Karin Rodrigues***	***System Operations. Control Room Operations***	***Proof Of Concept / Prototype***

Scope and Objectives of Use Case

This section outlines the objective and the scope of the use case

Related business case	AEMO: Operations Technology Programme. Operations Technology Roadmap, G-PST CROF Vision
Objective	A. Develop a use case methodology, risk assessment and data assessment for the applying of AI/ML to system operations and control room developments. B. Develop a test use case with the methodology using synthetic data and real data (from AEMO) to prove or refine the efficacy of the methodology.

Scope	Use synthetic and real alarm data to develop the use case on alarm data and event analysis. Consider 4 elements of the use case, each with a different but suitable algorithm and methodology: Alarm Forecasting Alarm Spike Detection Alarm Chatter Detection Alarm Pattern Matching

Narrative of Use Case

This section has a more complete description of the use case

Narrative of Use Case	*The aim of the use case is to make alarm management more intelligent and usable for the human operator at the desk and to reduce the burden of alarm overload on operators especially during alarm "spikes". Machine learning is a promising application given that alarms are time series datapoints with well structured information schemas. But there are different needs for an alarm improvement effort including:* *The ability to forecast if alarm loading will increase and if there is an uptick in alarm activity in a particular area or combination of substations.* *The ability to identify what is happening in alarm "spikes" to clearly see what has occurred and what assets are involved.* *The ability to identify and suppress chattering alarms so that they do not clutter the alarm field* *The ability to match patterns in alarm datasets based on the words or alarm fields.*
Short description – max 3 sentences	*To explore and develop a use case for machine learning applications with alarm management and event analysis using a variety of techniques to assess efficacy. The aim is to make the control room desk operators job easier when analysing events and for day-day activities.*

General Remarks

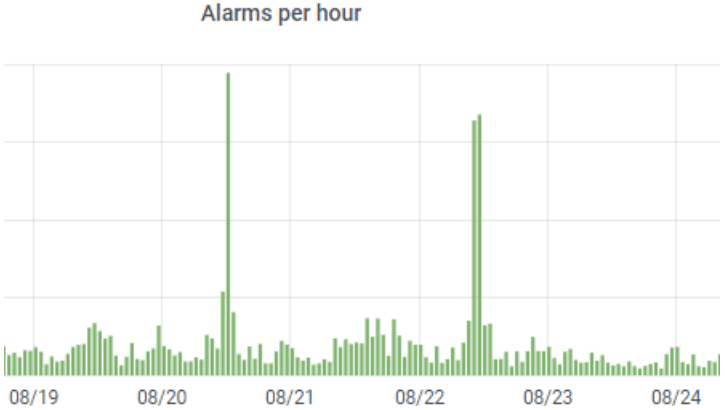
If there are any general remarks, they can be included in this section

General Remarks	*The algorithms can be trained and tested on both synthetic datasets and real datasets, without much by way of contextual additions. The algorithms should be generalisable to other control room and operational contexts if possible.* *Alarms in transmission and distribution operation control rooms generally are processed by large systems called EMS or DMS. Any use*

	case developed would eb outside of the EMS/DMS tied to historian databases and not fully integrated with EMS/DMS but the user interface esign of any system should match the general user interface design approach for operations technology.

Diagrams of Use Case

Diagrams helpful to the use case description can be drawn here. This may not be necessary in all use cases.

Diagram of Use Case	 *Example of alarm spiking behaviour which is typical of alarm load. This is the key part of the use case, to be able to identify alarm spikes, to eliminate or filter or supress chattering behaviour and to analyse the information in the alarm spike*.

Technical Details

Risks, Security and Trust

What are the risks and how will it be managed? The National Institute of Standards and Technology (NIST) AI Risk Management framework is the world leading AI risk management framework [8]. It has broad applicability to energy sector use cases and can be applied as a test case for the MLUCM. Not all points will be applicable for all use cases, but comments can be added to explain the context specific information.

[8] NIST AI Risk Management Framework Link: https://www.nist.gov/itl/ai-risk-management-framework

Risk Framework: NIST AI Risk Management Framework V2.0		
Map	MAP 1.1: Intended purpose, prospective settings in which the AI system will be deployed, the specific set or types of users along with their expectations and impacts of system use are understood and documented. Assumptions and related limitations about AI system purpose and use are enumerated, documented, and tied to (Test Evaluation, Verification, Validation) TEVV considerations and system metrics.	*Purpose: To design an intelligent alarm management and event analysis system.* *Deployment: On test servers, not in production environment for early-stage applications.* *Users: Data scientists initially to develop the system but ultimately the operators will use it to derive results and actionable information from the alarm datapoints.* *Assumptions: This will not be a perfect system and gaps and mistakes in forecasts will certainly happen, so full trust is not expected, it should work in parallel with the operator who use it as a decision support. All the underlying alarm data will be available to the operator to view as required.* *The Test Evaluation Validation Verification Cycle will be implemented during the development phase of the project on synthetic and real datasets.*
	MAP 1.2: Inter-disciplinary AI actors, competencies, skills, and capacities for establishing context reflect demographic diversity and broad domain and user experience expertise, and their participation is documented. Opportunities for interdisciplinary collaboration are prioritized.	*The team consists of PHD researchers in the AI ML field to develop the prototype. The team will be guided by engineering subject matter expertise in Australian network operations and intended operators of the developed system. Data scientists will test the prototype and comment on the results. There is broad diversity in the collaborative and there are no single points of failure.*
	MAP 1.3: The business value or context of business use has been clearly defined or – in the case of assessing existing AI systems – re-evaluated.	*Value: Reduce data overload to operators. Allow them to spend time risk assessing the power system rather than parsing through thousands of alarms during major incidents. Supply may be restored faster with smarter alarm systems which improve decision support. Utilising the vast quantities of alarm and system data available is also of value.*
	MAP 1.4: The organization's mission and relevant goals for the AI technology are understood.	*The use case aligns with CSIRO, EPRI, RMIT AEMO vision for excellence and innovative applications of the latest science and engineering technology.*
	MAP 1.5: Organizational risk tolerances are determined.	*Beyond data breaches which have major controls, there are no risks to the AI system for such early-stage development. This will eb reassessed for future development as the solution becomes more mature.*
	MAP 1.6: Practices and personnel for design activities enable regular engagement with stakeholders and integrate actionable user and community feedback about unanticipated negative impacts.	*Weekly team meetings are held to discuss and develop the project. Twice weekly meetings with the steering group are also held to report progress and assess risks.*

Risk Framework: NIST AI Risk Management Framework V2.0		
	MAP 1.7: System requirements (e.g., "the system shall respect the privacy of its users") are elicited and understood from stakeholders. Design decisions take socio-technical implications into account to address AI risks.	*Requirements are defined by the project team and end users. In early stage prototyping the requirements are defined during further development phases.*
	MAP 2.1: The specific task, and methods used to implement the task, that the AI system will support is defined (e.g., classifiers, generative models, recommenders).	*Forecast – Time series forecasting* *Spike Detection – Supervised machine learning* *Chatter Detection – Supervised machine learning* *Pattern Matching - Clustering*
	MAP 2.2: Information is documented about the system's knowledge limits and how output will be utilized and overseen by humans.	*This will be far from a perfect system, its accuracy will be continually tested and tuned over time to try to improve accuracy, but it will not be 100% accurate. Fall back to the core time series alarm information will always be available to the operators.*
	MAP 2.3: Scientific integrity and TEVV considerations are identified and documented, including those related to experimental design, data collection and selection (e.g., availability, representativeness, suitability), and construct validation.	*Training/Testing will be split 80/20 unless specified on datasets.* *The output of the system is clearly identified and labelled for evaluation against previous outputs.* *Verification and Validation will be carried out by the subject matter experts who analyse the results, provide feedback for the model t be tuned.*
	MAP 3.1: Benefits of intended system functionality and performance are examined and documented.	*Benefits are clear and are regularly evaluated with SMEs in operations.*
	MAP 3.2: Potential costs, including non-monetary costs, which result from expected or realized errors or system performance are examined and documented.	*Limited computation costs but these can be quantified. Costs of misdiagnosis are high and so operators are encouraged that this is a co-pilot design for decision support only.*
	MAP 3.3: Targeted application scope is specified, narrowed, and documented based on established context and AI system classification.	*Should be applied in test environment for initial phase, may be in simulator for evaluation by operators. Early-stage development should not be deployed in production environments or on real time systems. This should be evaluated later.*
	MAP 4.1: Approaches for mapping third-party technology risks are in place and documented.	*Open-source python packages are used in development such as tensor flow. These are safe but it should be documented what packages are needed and maintained for development training and development.*
	MAP 4.2: Internal risk controls for third-party technology risks are in place and documented.	*Controls in place within the project team*
	MAP 5.1: Potential positive and negative impacts to individuals, groups, communities, organizations, and society are regularly identified and documented.	*Not relevant in this context, beyond societal impacts of blackouts which will not be in the control of operators, who will be responding to incidents and not creating new incidents.*
	MAP 5.2: Likelihood and magnitude of each identified impact based on expected use, past uses of AI systems in similar	*N/A*

Risk Framework: NIST AI Risk Management Framework V2.0		
	contexts, public incident reports, stakeholder feedback, or other data are identified and documented.	
	MAP 5.3: Assessments of benefits versus impacts are based on analyses of impact, magnitude, and likelihood of risk.	*Risks are limited for early-stage development. This will be re-assessed at a later stage of development.*
Measure	MEASURE 1.1: Approaches and metrics for quantitative or qualitative measurement of the most significant risks, identified by the outcome of the Map function, including context-relevant measures of trustworthiness are identified and selected for implementation. The risks or trustworthiness characteristics that will not be measured are properly documented.	*Trustworthiness metrics are based on the accuracy of the performance and the risk of not identified spikes which will also be reported. There are risks with over trusting these types of applications which are documented, and stakeholders are informed.*
	MEASURE 1.2: Appropriateness of metrics and effectiveness of existing controls is regularly assessed and updated.	*Yes, the risks particularly with accuracy metrics will be continually assessed and adjusted.*
	MEASURE 1.3: Internal experts who did not serve as front-line developers for the system and/or independent assessors are involved in regular assessments and updates. Domain experts, users, and external stakeholders and affected communities are consulted in support of assessments.	*System operators and System operations managers and SMEs are involved in development.*
	MEASURE 2.1: Test sets, metrics, and details about the tools used during test, evaluation, validation, and verification (TEVV) are documented.	*Yes, the test, training and synthetic development data are available and the information about the data schema and context is developed.*
	MEASURE 2.2: Evaluations involving human subjects comply with human subject protection requirements; and human subjects or datasets are representative of the intended population.	*N/A – one operator is used, but this role can be shared among the operator cohort.*
	MEASURE 2.3: System performance or assurance criteria are measured qualitatively or quantitatively and demonstrated for conditions like deployment setting(s). Measures are documented.	*There are accuracy and performance metrics for all elements of the use case.*
	MEASURE 2.4: Deployed product is demonstrated to be valid and reliable. Limitations of the generalizability beyond the conditions under which the technology was developed are documented.	*The prototype is generalisable based on a defined input schema which can be adjusted as required. Valid and reliable on the real AEMO datasets.*
	MEASURE 2.5: AI system is evaluated regularly for safety. Deployed product is demonstrated to be safe and can fail	*Accuracy and efficacy are judged by the project team and safety risks assessed.*

Risk Framework: NIST AI Risk Management Framework V2.0		
	safely and gracefully if it is made to operate beyond its knowledge limits. Safety metrics implicate system reliability and robustness, real-time monitoring, and response times for AI system failures.	
	MEASURE 2.6: Computational bias is evaluated regularly, and results are documented.	*Early-stage prototyping but this is considered by the project team and developers as required.*
	MEASURE 2.7: AI system resilience and security is evaluated regularly and documented.	*Early-stage prototyping but this is considered by the project team and developers as required.*
	MEASURE 2.8: AI model is explained, validated, and documented. AI system output is interpreted within its context and to inform responsible use and governance.	*Clearly explained in the methodology section for Scope and Objectives. The users will eb aware of the objectives and scope. .*
	MEASURE 2.9: Privacy risk of the AI system is examined regularly and documented	*The underlying data is secured operational data and is sensitive, the tools are deployed on the servers of the system operator and data is shared, only when contractual obligations are met.*
	MEASURE 2.10: Environmental impact and sustainability of model training and management activities are assessed and documented	*Not assessed for early-stage development*
	MEASURE 3.1: Approaches, personnel, and documentation are in place to regularly identify and track existing and emergent risks based on factors such as intended and actual performance in deployed contexts.	*Early-stage development so no formal automated reporting system exists but risks are assessed as part of regular project meetings.*
	MEASURE 3.2: Risk tracking approaches are considered for settings where risks are difficult to assess using currently available measurement techniques or are not yet available.	*Early-stage development so no formal automated reporting system exists but risks are assessed as part of regular project meetings.*
	MEASURE 4.1: Measurement approaches for identifying risks are connected to deployment context(s) and informed through consultation with domain experts and other end users. Approaches are documented.	*The domain experts inform the development process and steer the team clear of risks where necessary.*
	MEASURE 4.2: Measurement results regarding system trustworthiness in deployment context(s) are informed by domain expert and other stakeholder feedback to validate whether the system is performing consistently as intended. Results are documented.	*The domain experts inform the development process and steer the team clear of risks where necessary.*
	MEASURE 4.3: Measurable performance improvements (e.g., participatory methods) based on stakeholder consultations are identified and documented.	*Performance is regularly traced for the different elements and benchmarks are measured and documented.*

Risk Framework: NIST AI Risk Management Framework V2.0		
Manage	MANAGE 1.1: Determination is made about whether the AI system achieves its intended purpose and stated objectives and should proceed in development or deployment.	*Not appropriate at this stage. Will be assessed after stage 3 of the project.*
	MANAGE 1.2: Treatment of documented risks is prioritized based on impact, likelihood, and available resources methods.	*Risks are regularly assessed by the project team.*
	MANAGE 1.3: Responses to the most significant risks, identified by the Map function, are developed, planned, and documented. Risk response options can include mitigating, transferring, sharing, avoiding, or accepting.	*Risks are regularly assessed by the project team.*
	MANAGE 2.1: Resources required to manage risks are considered, along with viable alternative systems, approaches, or methods, and related reduction in severity of impact or likelihood of each potential action.	*Risks are regularly assessed by the project team.*
	MANAGE 2.2: Mechanisms are in place and applied to sustain the value of deployed AI systems.	*Early-stage development, this will be more applicable to future stages of the system*
	MANAGE 2.3: Mechanisms are in place and applied to supersede, disengage, or deactivate AI systems that demonstrate performance or outcomes inconsistent with intended use.	*Early-stage development, this will be more applicable to future stages of the system*
	MANAGE 3.1: Risks from third-party resources are regularly monitored, and risk controls are applied and documented.	*Risks with third party or open-source software issues are managed by the project team.*
	MANAGE 4.1: Post-deployment system monitoring plans are implemented, including mechanisms for capturing and evaluating user and stakeholder feedback, appeal and override, decommissioning, incident response, and change management.	*Early-stage development, this will be more applicable to future stages of the system*
	MANAGE 4.2: Measurable continuous improvement activities are integrated into system updates and include regular stakeholder engagement	*Early-stage development, this will be more applicable to future stages of the system, however the project team engages with a team of SMEs and domain experts.*

Equivalent Use Cases

Details of how do other equivalent entities solve the problem, requiring use case development?

Equivalent Use Cases	
Neighbouring utility / regional level stakeholder use case application	*There are no known applications of a similar AIML alarm management system anywhere in Australia.*
Country level stakeholder equivalent applications	*There are no known applications of a similar AIML alarm management system anywhere in Australia.*
Global level stakeholder equivalent applications	*There may be some applications but the are limited to academic treatments and rarely deployed in real time operations. The use case will be using novel approaches.*
Academic and research review	*Some*
Opportunities for collaboration with stakeholders locally, regionally	*Yes, the system should be generalisable in other operations and control room contexts so can be applied to other TNSPs as part of other future projects.*
Academic and research collaboration explored?	*The team partners with RMIT and EPRI and CSIRO with AEMO the system operator. This gives broad academic and research institute experience.*

Use Case Evaluation Methods

Details as to why AIML is being considered for this use case.

Evaluation	
Why is current practice insufficient/inefficient?	*Data overload, time spent filtering, sorting, zooming and scrolling to find key and important alarm information among hundreds of noisy alarm indications.*
Why is machine learning being considered?	*Statistical analytical approaches may be too computationally intensive. Time series forecast and predictions show good early stage promise for alarm management use cases.*
What alternatives were considered and why were they discounted?	*Statistical analysis using averaging, standard deviation and rule-based approaches such as > alarms in 1 minute = chattering. These are slow and computationally intensive. Machine learning should be faster.*

Data Standards

What data is being used in the use case, what is the data quality and is there a data standard?

Data Standards	
What data is required?	*Alarm time series data*
Data Source (Database)	*SCADA / EMS alarm database*
Data type (Time Series/Image/Text)	*Time series*
Data Standard (CIM, 61850.)	*None – standard alarm database schema from the vendor product*
Data security levels (critical energy, personal data)	*Sensitive critical energy sector data. Must not be released to public*
Structured / Unstructured data	*Semi structured the alarm data is in a structured database with defined fields but the text in the alarm message can eb unstructured with no discernible patterns.*
Data Quality (Low/Medium/High)	*Medium, may be parsed and improved with pre processing and cleansing.*
Data Cleaning Time	*Medium*
Data Housing (Location – on shore, on shore, cloud etc)	*On prem servers for real testing. Servers can be production or test environments depending on the use cases.*
Potential Inclusion of PII (Personal Identifiable Information)	*There will be no PII in the datasets and the nature of the use case and algorithm will not be concerned with PII identification.*
Assessment of Potential for Data Biases in Models	*There may be biases developed and overfitting of the models if only used on AEMO data. Model may not be generalisable but should be tested on other relevant datasets.*

Baselines and Benchmarks

How is success measured and evaluated over time and what is the latest status.

Baselines and Benchmarks	
What metric will be used for evaluation model performance (speed, accuracy)?	*Alarm load – alarms per minute with moving average window of 30 minutes. Accuracy of each element defined by* *Forecasting – RMSE, MAE* *Spike Detection: Accuracy and false positive and negative accuracy and speed of processing* *Chatter detection - Accuracy and false positive and negative accuracy and speed of processing* *Pattern Matching – Efficacy of identified clusters.*
How will the baseline metric be measured?	*Statistical analysis of the dataset.*
What is the baseline value for the metric?	*Alarms per minute*
Current benchmark	*Dependent on dataset. Approx 500 per minute.*
Benchmark Algorithm version (with date)	*2023 June 18*
Benchmark algorithm comment	*Developed algorithms for stage 2 project*
Industry standard benchmark?	*N/A none to date.*
Explainability of the Model (Does the new model improve the capability make decisions)	*It should limit superfluous information and noisy data to only limit the alarm data to actionable information within context. This should help with explainability of the model.*
Can inputs be traced back to outputs?	*In some algorithms yes but depending on the model used and algorithm such as deep neural networks this may need to be assessed through the life cycle.*

Compute Resources

How is success measured and evaluated over time and what is the latest status.

Compute Resources	
What compute resources are expected to be used?	*Machines for development. Test servers for deployment.*
CPU or GPU?	*CPU for first stage, may be an increase needed over time.*
Cloud or on Prem Resources?	*Prem for earl-stage*
Cloud service provider?	*N/A*
Why is cloud used or not used?	*Security of data and deployment of the tool for a control centre application requires on-prem. May be re-assessed for future iterations.*
Expected Cost of compute resources?	*Negligible for early stage. May increase over time/*
Integrated with business systems?	*Early-stage not integrated but may be integrated in stage 2.*

Human Resource Input

Who will be working on this use case development

Human Resources	
Exec / Manager Sponsor	*Chris Knight, Tjaart van der Walt, Luke Robinson,*
Product Manager	*Adrian Kelly*
Developer(s)	*Xinghuo Yu, Chen Liu, Geordie Dalzell*
Business Subject Matter Expert (engineering)	*Karin Rodrigues, Elena Kranz*
External resources for development and other expertise	*Mahathir Almashor (CSIRO)*

External Vendor Collaboration

How will software vendors and developers collaborate or develop this use case?

External Vendor Collaboration	
What external vendors will collaborate on the project	*N/A*
What role will they play	*N/A*
Where have they demonstrated previous capability associated with, he uses case	*N/A*
How will they access and securely hold data ***What is the expected outcome***	*N/A*
For Open Source tools – licensing arrangements	*All open source tools have relevant open licences for use in the development of the use case.*