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

CSIRO GPST Research Stage 3

Adrian Kelly Senior Principal Technical Leader EPRI Europe

30th April 2024

RMIT UNIVERSITY AEMO AUSTRALIAN ENERGY MARKET OPERATOR EPRI

Citation

Global Power System Transformation Research Topic 3 - Control Room of the Future, 2024. CSIRO, Australia.

Copyright

Important disclaimer

Contents

Figures

Tables

Acknowledgments

1　Introduction

In 2021 CSIRO launched the Global Power System Transformation (G-PST) Research initiative to drive research innovation in the Australian electricity sector, structured around nine topics. EPRI have partnered with CSIRO to develop the G-PST Topic 3 Control Room of the Future (CROF) research roadmap, and since 2021 have been developing research to advance innovation on this key topic. The original 2021 roadmap[1] outlined the innovation pathways and actions needed, structured around the five key CROF pillars:

- Data
- Architecture
- Software Applications,
- Human Factors Operator Interactions
- Facilities and Equipment

The CSIRO CROF stage 2 work in 2022-3 began to work through the elements of the roadmap with focus on the CROF research pillars for data and software applications. The initial aims of the stage 2 research were to initiate and work on the core capabilities of artificial intelligence and machine learning (AI/ML) for real time operations applications, given the long gestational period of development for these technologies, it was important to start the research early and iterate through applications and work on real data. This stage 2 work involved close interaction between EPRI, Royal Melbourne Institute of Technology (RMIT) and AEMO to identify a methodology for developing machine learning projects, data and use cases. The project developed further to develop proof of concept prototypes applications for use on real AEMO data. The stage 2 project was completed in 2023 and the report on the project is available publicly on the CSIRO website[2].

---

[1] Topic 3 Control Room of the Future Research Roadmap *https://www.csiro.au/en/research/technology-space/energy/g-pst-research-roadmap).*

[2] Topic 3 Control Room of the Future Stage 2 Report: https://www.csiro.au/-/media/EF/Files/GPST-Roadmap/Final-Reports/Topic-3-GPST-Stage-2.pdf
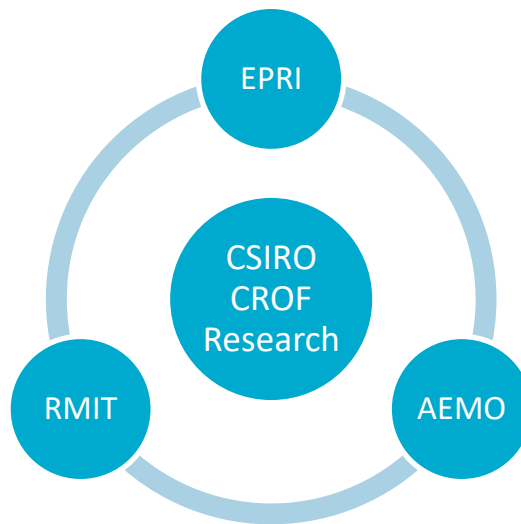
Figure 1 Project Team for Stage 2 and Stage 3 developments of the Topic 3 CROF research

In addition to the 2021 research roadmap and stage 2 research; CSIRO partnered with AEMO and EPRI, in 2022 to develop a targeted Operational Technology Roadmap for AEMO[3]. This report leverages the CSIRO research roadmap but with focus on the pathways for operational applications and technology developments in the coming decade, to meet the monitoring and assessment needs of AEMO in their role as the electricity system and market operator.

Within the 11 operational applications in the OTR – there was a focus on how to leverage emerging (AI/ML) technologies and systems, while emphasising the needs for iterative design over long time horizons, to mitigate trust and safety issues. See Figure 2 for the AEMO EMS/SCADA operations technology roadmap – which shows that AI/ML research is continuous, long-term process rather than a complete solution that can be deployed out of the box.

AI/ML development proceeds in parallel with other operations technology (OT) software application developments and deployments, while also retaining the ability to integrate and share data between the applications.

---

[3] AEMO, CSIRO – Operational Technology Roadmap Report 2022 https://aemo.com.au/-/media/files/initiatives/operations-technology-roadmap/executive-summary-report-for-the-otr.pdf?la=en

A Roadmap for Development through 2030+
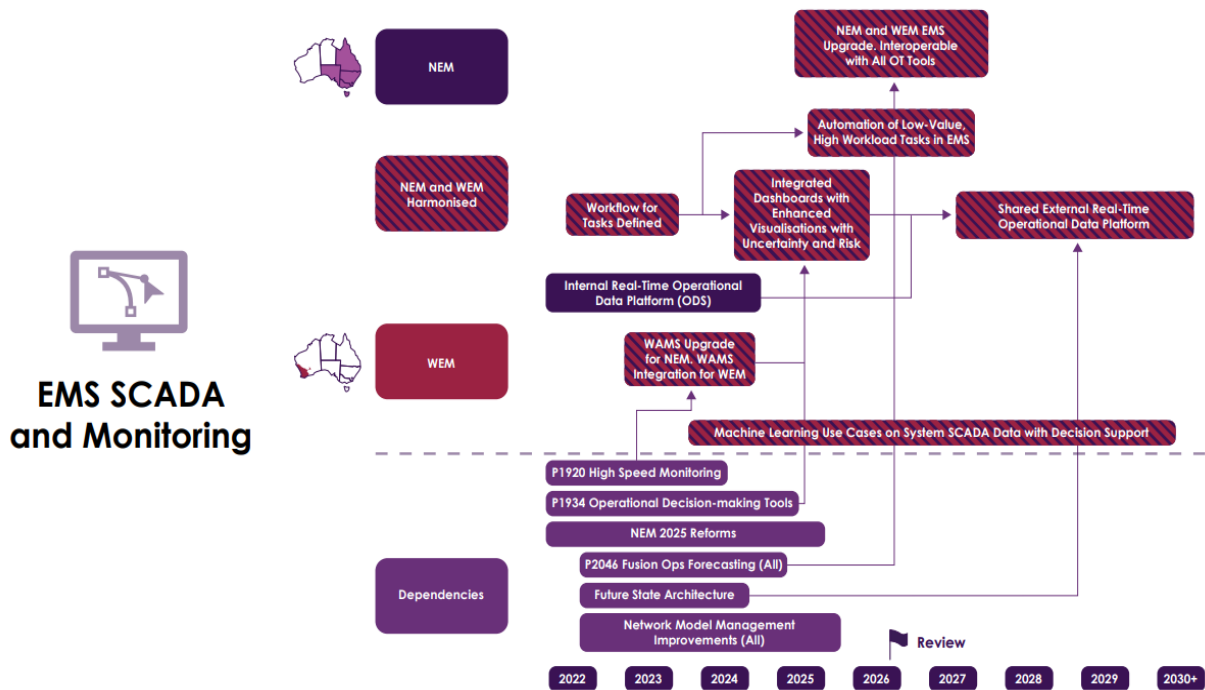
Figure 2 Extract from the AEMO/CSIRO OTR from 2022 roadmap for data and models which highlights the need for continuous development of AI/ML applications for electricity system control.

Since the launch of the 2021 roadmap and the OTR there has been era-defining transformational developments in AI/ML – characterised by the advent of large language models (LLM) and generative pre-trained transformers (GPT). These wider societal developments merit further exploration in the operational technology and data application context. This is primarily because of the large quantities of data that can be leveraged for training of AI/ML. This data is primarily text and numerical time-series data which is considered a good candidate for AI/ML use cases generally. Additionally, system operators in Australia and around the world are being faced with an increase in data from transmission resources and assets and would benefit greatly from the streamlining of information and process automation that AI/ML can be used for.

The framework and proof of concept development work from stage 2 was continued in the stage 3 project in 2023-4. EPRI are again partnering with RMIT and AEMO to continue the research. The focus remains on the data and applications pillar of the original 2021 roadmap and the AEMO OTR, but with an additional focus on the use of large language models in the operational context.

While research and innovation on time series operational data and alarms are high priorities for network operators, there was also a need, in stage 3 to begin steps on the roadmap focussed on operational modelling, given the broad ambition and vision of the original roadmap. The focus in stage 3 was on the need for streamlined validation processes for the operational model of the transmission network, including generation resources. EPRI partnered with AEMO – as the entity with responsibility for dynamic simulations – to begin work in this area with the development of a methodology for operational or real time model validation, that builds on current AEMO innovations in this area.

2   Overview of the Project

The scope of the project was structured around two pillars of the CROF roadmap and three tasks - detailed below in Figure 3.
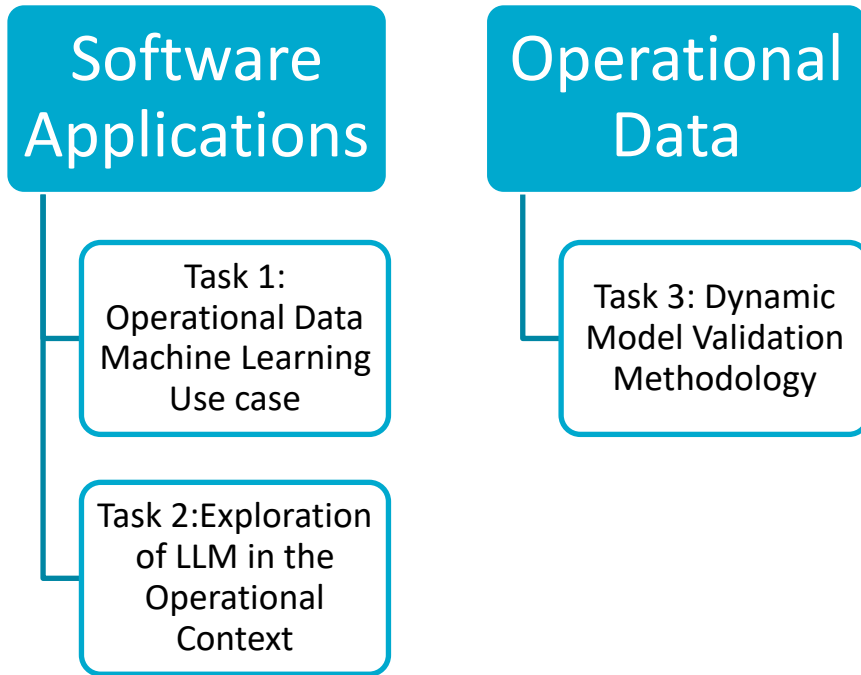
Figure 3 Structure of the CROF Stage 2 project

2.1   Task 1. Operational Data Machine Learning Applications

2.1.1   The need for development and innovation

Electricity network operators, in Australia and globally, regardless of their function (TNSP, MO, TO, TSO, ISO, DNSP, DSO) all have common features:

1. The need for operators in real time to process and act on a large quantity of real time data.
2. The growth of this data due to new generation, network technology, markets and interactions with neighboring or interconnected network operators.
3. The rapidly changing nature of the system that operators are monitoring and controlling due to decarbonisation and electrification.
4. The turnover in knowledgeable, experienced operational staff and difficulties replacing, retaining and training new operators.

While the quantity of data and the risks to networks are growing, the number of operators in control rooms is expected to stay relatively constant, the alarm data handling mechanisms in EMS/SCADA are not expected to evolve significantly in the near term. One way to redress the imbalance of increased data with finite human resources is to develop innovations in how data is processed, filtered and presented to operators in real time.

2.1.2   The Three Modes of Operator Cognition

Operators (and people more generally) can be considered to have three modes of cognition when faced with any type of problem to solve.

Figure 4 Illustrative example of the three modes of operator cognition, when trying to solve a problem.

These are:

- Sense Making – Characterised in Figure 4 as the yellow mode - where operators monitor data in real time and make sense of data and information. If an abnormal event occurs – in the sense making mode - operators identify the problem and diagnose the root cause.
- Decision Making – Characterised in Figure 4 as the Orange Mode – where operators build on their mental model constructed in the sense making mode and decide a course of action to solve the problem. Here operators leverage their intuition, experience and knowledge to determine the optimal solution. They can also leverage OT applications such as forecasts, simulation, optimisation, and risk assessment to get the decision.
- Action Making – Characterised in Figure 4 as the red mode above – once a decision is made in the orange mode the operator must execute an action and review the response. Actions can be switching events, dispatch of resources – primarily but can also include actions such as phone calls to dispatch personnel, or other escalations, reporting and logging. The operator should also continue to monitor the network post action.

The increase in data has resulted generally in data overload in control rooms, this makes the "sense making" mode more difficult as operators struggle to identify what has happened and why it has happened. In normal conditions it is also difficult for operators to determine if the network is behaving abnormally, due to its dynamic nature.

2.1.3   Vision for Development of Enhanced Sense Making Capability

In an idealised control room: a smart, real-time operational or alarm system would synthesise all operational data into digestible information for the operator to inform them of abnormalities in the system state. When a network disturbance occurs, it would identify the abnormality with

reference to past events and suggest what the likely cause to the system issue is with a potential for proposed actions.

The smart alarm system would be easily searchable and be able to return insights on the operational data based on clearly defined prompts.

The vision is summarised in Figure 4 and the research activities in this project are structured around delivering this vision. Machine learning is one of the key tools to achieving this vision.

<table>
<tr>
<td>Ability to summarise operational data and identify root causes of abnormalities</td>
<td>Ability to be efficiently prompted for insights into the data via text or voice</td>
</tr>
</table>

Figure 5 Summary of idealised alarm and operational data system

2.1.4   Delivering the Vision

To deliver the vision, using AI/ML - there are several important inputs to be considered:

- An extensive quantity of well-structured real operational data for training

- A variety of operational data sources, such as planned outage information, market notices, unplanned event and disturbance logs. All the data sets should be time synchronised and machine readable (text or numerical).

- The ability to label important data points in the operational datasets for supervised learning and the ability to train an algorithm in parallel with an expert operator, so that when presented with new information the algorithm can reference archive events with similar features and resultant actions.

These pre-requisites were achieved in partnership with AEMO.

The key value-add and research force multiplier with this project is that the prototype is designed to be applicable and useful in other operational contexts and should not be limited to the operational context of AEMO or transmission networks more broadly.
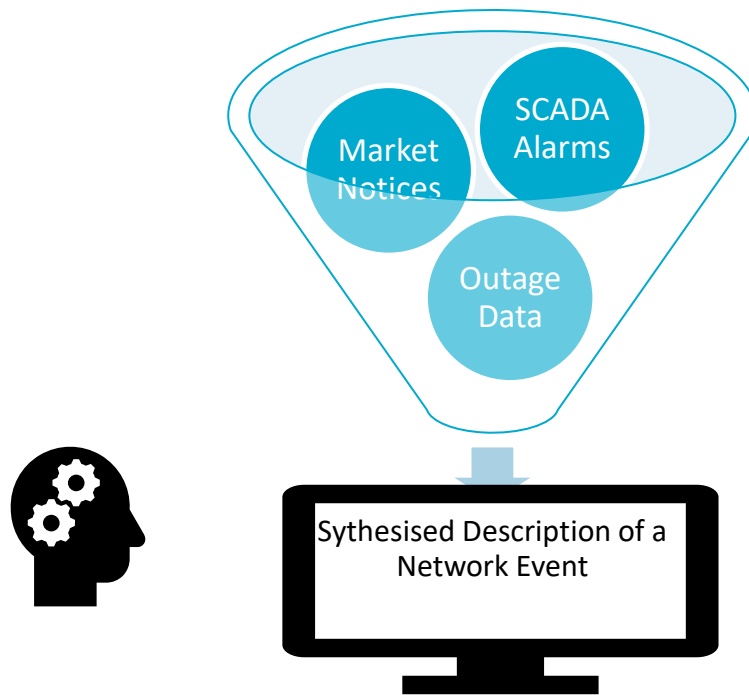
Figure 6 Plan for algorithm to synthesise operational data for human operator to process.

2.1.5 Scope and Deliverables

The work in Task 1 involves continuation of the stage 2 R&D – but with algorithms directly trained and applied on real AEMO operational data from diverse range of operational datasets. (The stage 2 work was primarily focussed on a synthetic dataset of operational data).

Ultimately the aim is for a deployment of an operational data prototype directly on AEMO systems that uses real time AEMO operational data, to augment operator sense making.

The ability to deploy the prototype is dependent on the maturity and security of the prototype and AEMO IT/OT policies.

2.2 Task 2 - Exploration of natural language and text-based machine learning and knowledge-based systems in system operations

2.2.1 The Need for Development and Innovation

The primary feature of operational data in operational technology applications is that they are **mostly text based and semi structured.** The data are classified as semi-structured as they generally have time stamps and the fields are consistently parameterised but, in some cases, the longer text description is unstructured, truncated and not of a consistent format – see example in the event_message filed in Figure 7.

Numerical, analogue information and data points and indicators trigger a text-based alarm in the alarm system when they breach the technical operational limit of the asset. For example:

ALPHA STATION LINE BETA 500 KV OVERLOAD 105 MVA LIMIT 100MVA

Asset and switchgear alarms are text-based such as an example of synthetic data shown in Figure 7.

EventId	EventTimeStamp	event_message	SCADA_Category	TOC	AOR	Priority_Code	Substation	DeviceType	Device
1	01/05/2017 00:00	Substation1317 STN STA_SERVICE AC SERV VOLTAGE NORMAL CHG-DET	MH	TMS	TOCMSDOC	3	Substation1317	STN	STA_SERVICE
2	01/05/2017 00:00	Substation1265 BLKCAR STUTR_PR1_UPLC CARRIER LEVEL LOW CHG-DET	ML	TMS	TOCPB	4	Substation1265	BLKCAR	STUTR_PR1_UPLC
3	01/05/2017 00:00	Substation1459 BLKCAR Device285 CARRIER LEVEL LOW CHG-DET	MH	TMS	TOCPB	3	Substation1459	BLKCAR	Device285
4	01/05/2017 00:00	Substation1459 BLKCAR Device285 SIGNAL RECEIVED NORMAL CHG-DET	MH	TMS	TOCPB	3	Substation1459	BLKCAR	Device285
5	01/05/2017 00:00	Substation275 FIBOPT Device4576 PRI #2 OPTIC/CH NORMAL CHG-DET	MH	TMS	TOCGR	3	Substation275	FIBOPT	LABAR_X_311L
6	01/05/2017 00:00	Substation749 RELAY L37_P2_SEL411L RELAY FAILURE NORMAL CHG-DET	MH	TMS	TOCBM	3	Substation749	RELAY	L37_P2_SEL411L
7	01/05/2017 00:00	Substation256 RELAY L532_SEL421313 SYSTEM NORMAL CHG-DET	MH	TMS	TOCBM	3	Substation256	RELAY	L532_SEL421313
8	01/05/2017 00:00	Substation113 RELAY L513_SEL411_P2 RELAY FAILURE NORMAL CHG-DET	MH	TMS	TOCPA	3	Substation113	RELAY	L513_SEL411_P2
9	01/05/2017 00:00	Substation460 RELAY L461 BKUP SYSTEM FAIL NORMAL CHG-DET	MH	TMS	TOCPA	3	Substation460	RELAY	L461
10	01/05/2017 00:00	Substation232 BLKCAR Device689 SIGNAL RECEIVED NORMAL CHG-DET	ML	TMS	TOCED	4	Substation232	BLKCAR	Device689
11	01/05/2017 00:00	Substation1322 RELAY Device1492 #1 PRIMARY RELAY NORMAL CHG-DET	MH	TMS	TOCGR	3	Substation1322	RELAY	Device1492
12	01/05/2017 00:00	Substation303 RELAY L532_SEL421 SYSTEM NORMAL CHG-DET	MH	TMS	TOCBM	3	Substation303	RELAY	L532_SEL421
13	01/05/2017 00:00	Substation611 BLKCAR L752_PULSAR SIGNAL RECEIVED BLOCK TRIP CHG-DET	ML	TMS	TOCBR	4	Substation611	BLKCAR	L752_PULSAR
14	01/05/2017 00:00	Substation78 BLKCAR Device753 SIGNAL RECEIVED TRIP ENABLED CHG-DET	ML	TMS	TOCED	4	Substation78	BLKCAR	Device753
15	01/05/2017 00:00	Substation1118 RELAY Device4574 SYSTEM NORMAL CHG-DET	MH	TMS	TOCBR	3	Substation1118	RELAY	FCPT_SEL421_5_

Figure 7 An example of the format of SCADA alarm text in operational technology systems such as EMS (Synthetic Data)

AEMO market notices, issued publicly and to market participants are also text based as shown below. Network Outage System (NOS) is similar – open, free to the public, text based and semi structured.

114853 RECLASSIFY CONTINGENCY 17/02/2024 09:34:56 PM

Cancellation of a Non-Credible Contingency Event: Tumoulin - Woree No.877 275 kV line and Chalumbin - Turkinje No.7166 132 kV line in QLD1 due to Lightning.

AEMO ELECTRICITY MARKET NOTICE

Cancellation of reclassification of a Non-Credible Contingency Event as a Credible Contingency Event due to Lightning. AEMO considers the simultaneous trip of the following circuits is no longer reasonably possible. There is no longer any lightning activity in the vicinity of the following lines. Accordingly its classification has reverted to a non-credible contingency event.

Region: QLD1

Lines: Tumoulin - Woree No.877 275 kV and Chalumbin - Turkinje No.7166 132 kV

Cancellation time: 17/02/2024 21:34

Constraint set(s) revoked:
NONE

Refer AEMO Electricity Market Notice 114851

Auto-generated on behalf of Manager NEM Real Time Operations

Figure 8 Example market notice published online on February 17th, 2024.

The other value of operational text is that it is structured in a consistent format for the most part. Text based data that is at least somewhat structured and which includes a more natural or conversational style are valuable for machine learning and large language model applications. So operational data sets are potentially very good candidates for further exploration with LLM.

Despite the good potential and the availability of good, text-based data, using LLM on operational data is a very new and emerging research area, given the transformative impact of LLM on wider society since the public release of Chat GPT in 2022.

2.2.2   Potential Use Cases for LLM in the Operational Context

The three modes of operator cognition described in Section 2.1.2 can potentially be augmented by LLMs.

Figure 9 The three modes of cognitive processing in the operational context

Sense Making Applications of LLM

LLMs can be used to help operators make sense of large quantities of text-based data, in particular the ability to search archive material using prompts could improve operator accuracy and efficiency when diagnosing issues. Some other potential use cases are listed below:

a) Finding patterns in operational and alarm text data.
b) Using filters and query creation using information in voice and text.
c) Search, summarisation, and citation from operational data points from different datasets.

Action Making Applications of LLM

When an operator makes sense of data and decide on a course of action – they must implement that action. These actions (red mode) can be controlling actions on the network such as breaker operations or generator re-dispatch or the actions could be the creation of a report or log, dispatching of personnel or the creation of a switching instruction.

Non-network control actions have powerful potential applications of LLMs, using natural language processing (NLP) techniques from a person's voice. There is a greater degree of information in a faster time frame contained within voice communications. LLMs can be leveraged for efficient actions, to reduce manual data entry and administrative activities and in the operational context than typing and data entry.

Some potential use cases in the action domain are:

a) Plan or switching instruction creation.

b) Reporting and log entry that involves operator typing.
c) Interaction with OT systems through voice commands.
d) Retrieval search of procedures/protocols

By using LLM in the operational context, the operator's efficiency can be maximised. The aim should be to spend less time on manual data entry, document search or administrative task to spend more time in the decision making and sense making modes, risk assessing and planning for upcoming operational activities.

An illustrative example of the hypothetical operational process to resolve an unplanned event process is shown in Table 1.

Table 1 Example of an operational process to resolve an unplanned event on the network that could be augmented by LLM/NLP

Manual Operation Process	Process Augmented by LLM/NLP
Event occurs operator searches alarm archive for past events. Must mentally filter out superfluous information to find the most relevant data based on their understanding.	Operator types or says "find last occurrence of this alarm"
Operator searches reporting log for prior occurrences	Operator types or says "find when an incident like this last occurred"
Operator must search to file system and document archive for procedure for how to handle the unplanned event	Operator types or says "how do I resolve the problem in this station, cite and give access to the sources"
Operator must reconfigure the network through switching actions, first manually creating a document switching order	Operator types or says, "Create switching plan to close circuit breaker AB and BC" It is presented to the operator for approval.
Operator must make a phone call to dispatch a field operator to the station to investigate and manually switch.	Operator types or says, "dispatch the next available operator to station ABC to investigate breaker AB opening" It is presented to the operator for approval.
Operator must manually type a description of the incident into the logging system for records.	Operator types or says a verbal description of the event and his command is parsed into the relevant operational database. It is presented to the operator for approval.

The decision-making (orange) mode is not well suited to LLM application, at this level of maturity as this mode relies on simulation, forecasting, risk assessment, optioneering and is the most knowledge intensive of the three modes. Given that LLMs rely on unsupervised learning and

2.2.3 Scope and Deliverables

Task 2 of this project involves exploration of the feasibility of possible use cases for large language models in the system operations domain. LLMs could be potentially useful and powerful in operations but to date, applications of LLM in operations control rooms are rare and these explorations are very novel. Given this is emerging technology - it was unclear at the outset what could be possible for training and deployment within the stringently secure operational technology environments.

2.3 Task 3 - Network and Generator Model Validation Processes

2.3.1 The Need for Development and Innovation

Traditionally dynamic simulations and the development and maintenance of the dynamic model of the system were in the domain of the network planning teams of system operators. When new assets or new generation connections were being planned – network planners would study the dynamic performance of the assets through a series of rigorous simulations against a series of tests such as long duration faults or under frequency events. In the Australian context, for 25 years one of the obligations for connecting to the NEM has required the provision of time domain power system models of the generating systems. These models were then also used by the developers to prove compliance with National Electricity Rules (NER) requirements. A process for validation of the dynamic models was introduced more than 20 years ago.

However, the process of validation was difficult to achieve practically and not systematically automated. Due to a lack of relevant large disturbance, actual validation is difficult. Validation was based on normal operation points and post large disturbances validation is laborious and reliant on high-speed recording.

As the network has evolved, the traditional generator models may not accurately reflect the reality of the asset and the pace of growth of inverter-based resources, especially embedded IBR and DER possess challenges for modelling and simulation and opens the network to risk.

In recent years dynamic simulation and security assessment has shifted from solely a network planning competency to a real time operations competency. The change in the network characteristics requires accurate real time simulations to allow operators to assess security. TSOs/ISOs study the network in real time, with dynamic security of:

- Voltage Stability
- Frequency Stability
- Transient Stability
- Small Signal Stability

In the coming years, according to the OTR, AEMO (as led by other TSO/ISO around the world) will enhance real time dynamic security assessment to include a look-ahead capability with near term

forecasts for demand, DER and VRE. There will likely also be a need to study converter driven instability with more granular simulations and models. These enhancements will allow proactive network and market management, but the model accuracy will be critical to decision making.

2.3.2 Dynamic Model Assumptions

Dynamic simulation studies, whether in the planning or operations domain have always relied on two important assumptions:

1. The underlying model of the system including all resources and assets is accurate and kept accurate through validation.
2. The models of newly connected asset, despite not being manufactured or in-service yet was accurate.

These assumptions held true, so long as dynamic stability was not a major network issue i.e., if the network was not being run close to the boundary of its operational envelope. Limit equations in NEM dispatch engine are determined by dynamic simulations but these equations have safety margins in built to allow for model inaccuracy.

In recent years all TSOs/ISOs are pushing their networks to the boundaries of their stable regions, to accommodate more smaller, variable, decentralised renewable generation resources.

The assumptions for dynamic stability simulations (globally) are being challenged in fundamental ways in recent years:

a) The original operations model may be separate from the planning dynamic model and may use different simulation applications. The operations model may have been baselined off the planning dynamic model, but in some jurisdictions without a common model and feedback loop for updates, there is a divergence between planning models and operational models.
b) Due to the massive increase in VRE and DER, it is very challenging to maintain, validate and correct a single, accurate dynamic model of the network. Multiple small, inverter-based resources are harder to model than single large conventional generator resources.
c) New resources are inverter based with power electronic controllers, with multiple parameters and control intricacies. The parameters of the model when commissioned may be different to the model presented in the planning time frame when it may be uncertain what actual technology type is being procured and installed, despite best efforts.
d) The models in some cases are black box, for commercial reasons and the OEM is not incentivised to openly release the models or parameters. The developers of VRE and DER systems do not have the modelling and simulation competency to address issues and rely on the OEM.

Note: The CSIRO G-PST research agenda includes Topic 2 Stability Tools & Methods which is being developed in parallel to Topic 3 CROF.

Dynamic model inaccuracy puts networks at risk as dynamic security assessments may not detect security issues that require mitigation controls in real time. It also has knock on economic impacts, as dynamic security assessments simulation results may trigger constraints in the market – so if the simulation is not accurate, even including the safety margins - the market may be unnecessarily constrained.

2.3.3   Model Validation Enablers

The one and only way to validate a dynamic model of a network is to compare simulation and models performance based on measured data from real events. In the past this was not technically possible as data available to TSO/ISOs (SCADA data) that was needed to validate performance was not of a fast enough granularity (frequency) to determine if the asset performed to meet its model performance in a simulation. High frequency high speed recording devices are required for validation. Rules and compliance regulations have required generators to install high speed monitoring equipment and meters and make it available to TSO/ISOs on request, but this was usually post-event and manual and laborious to gather and assess.

The ability to simulate based on the realistic network conditions was also difficult as reconstructing an event meant an accurate representation of all assets at the time of the fault.

In recent years both technical constraints on model validation have been reduced. There has been a slow growth and proliferation of high-speed monitoring (HSM) devices (Phasor Measurement Units and Digital Fault Recorders) on transmission networks. The cost of the devices has reduced and the systems to manage the data from them are also now very mature. New connection on the transmission network is mandated to install a HSM at their bulk system connection point and make it available to the TNSP. There is ongoing work to facilitate the sharing of HSM data between the TNSPs and AEMO for real time monitoring.
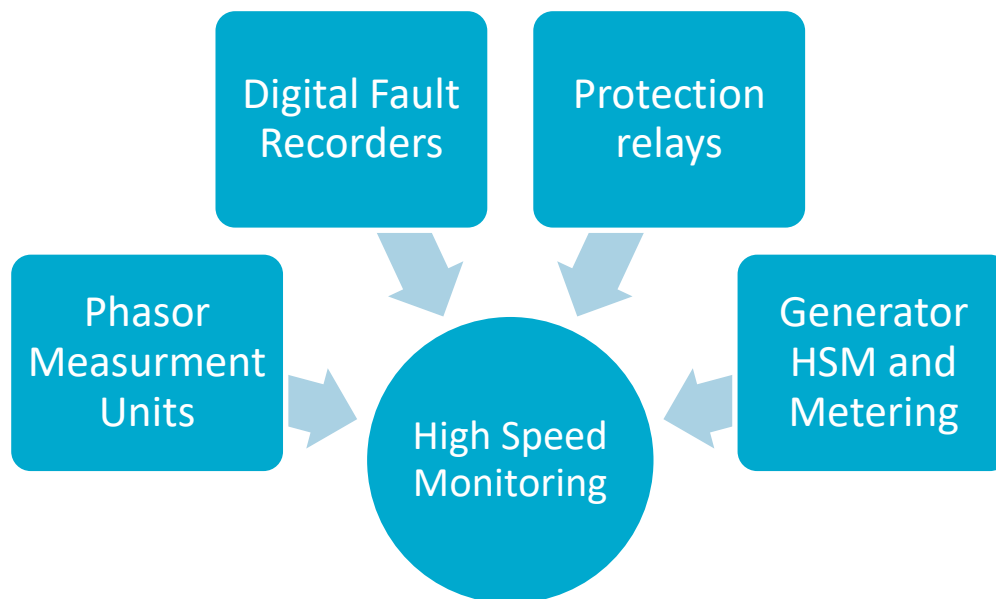
Figure 10 The categorisation of HSM devices

In addition, state estimation and enhanced simulation technology has made it easier to take historical snapshots of the network to replicate realistic conditions for major events in simulations.

2.3.4   Interconnecting Modelling Entities

To achieve an accurate, usable, automated model validation process there are several large interconnecting functions and asset owners with key interest in curating and maintaining the

models. These are illustrated and described in Figure 11. Any automated system will require collaboration and data sharing between the different entities.

ISO (AEMO)
- Maintain the dynamic models
- Perform the dynamic simulations
- Owner of the operational data

Generation Asset Owner
- Owner of asset
- Providor and updater of model
- Owner of HSM data on their assets for compliance

Transmission Asset Owner
- Owner of PMU, DFR, Relay
- Owner of model collection process
- Pass disturbance data to AEMO
- Connected with DNSPs
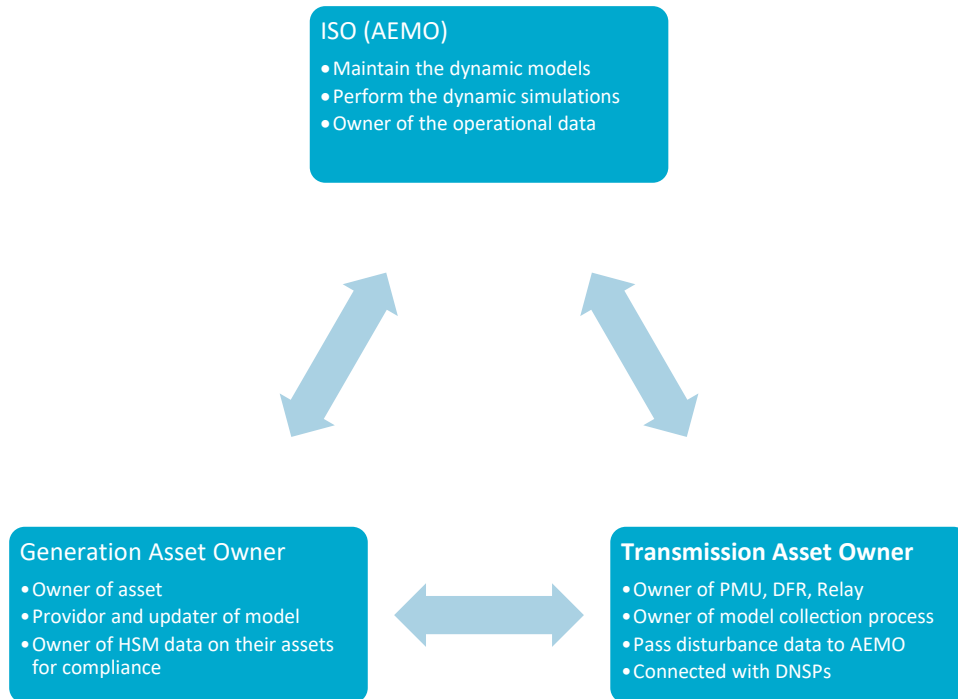
Figure 11 Illustration of interdependencies of the processes and data

2.3.5   Vision for Development of an Operational Digital Twin

The innovation in Task 3 is to work with AEMO to develop a methodology for the validation of dynamic model performance using real time operational data such as HSM.

The long-term vision is to have an operational "digital twin" of the transmission network.

A digital twin is an automated, continuously running validation process that links HSM data to the dynamic simulations and models that are parameterised based on the real time network. The validation process would automatically identify model anomalies based on real power system disturbance events and data and suggest model changes based on machine learning that could be communicated to the asset owner and OEM for further information and changes to be made.
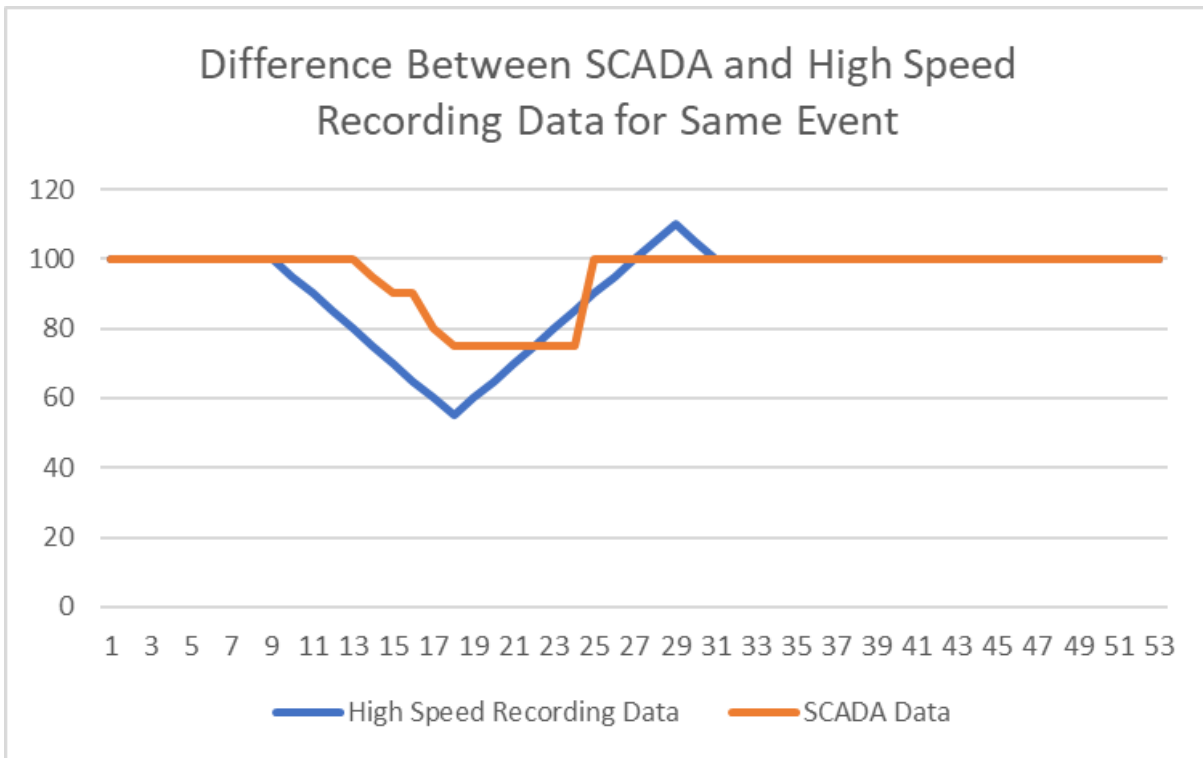
Figure 12 Illustrative example of difference between low frequency SCADA data and high frequency HSM data.

2.3.6 Scope and Deliverables

This task involves engaging with modelling subject matter experts in AEMO to assess current model validation processes and activities and to define a methodology for automatically validating dynamic models for use in dynamic simulations using high-speed data recorders.

3   Research Results

3.1   Task 1 – Operational Data Machine Learning Applications

3.1.1   Task Outcome

A prototype ML based system to assess alarms link to similar occurrences from the archive was developed. This is operational on real AEMO operational data and on AEMO systems. It is not deployed in real time operations yet. A sustainable framework for improvement of the prototype over time was also created so that performance can be improved, and new features added. The process for developing the process is explained in this section.
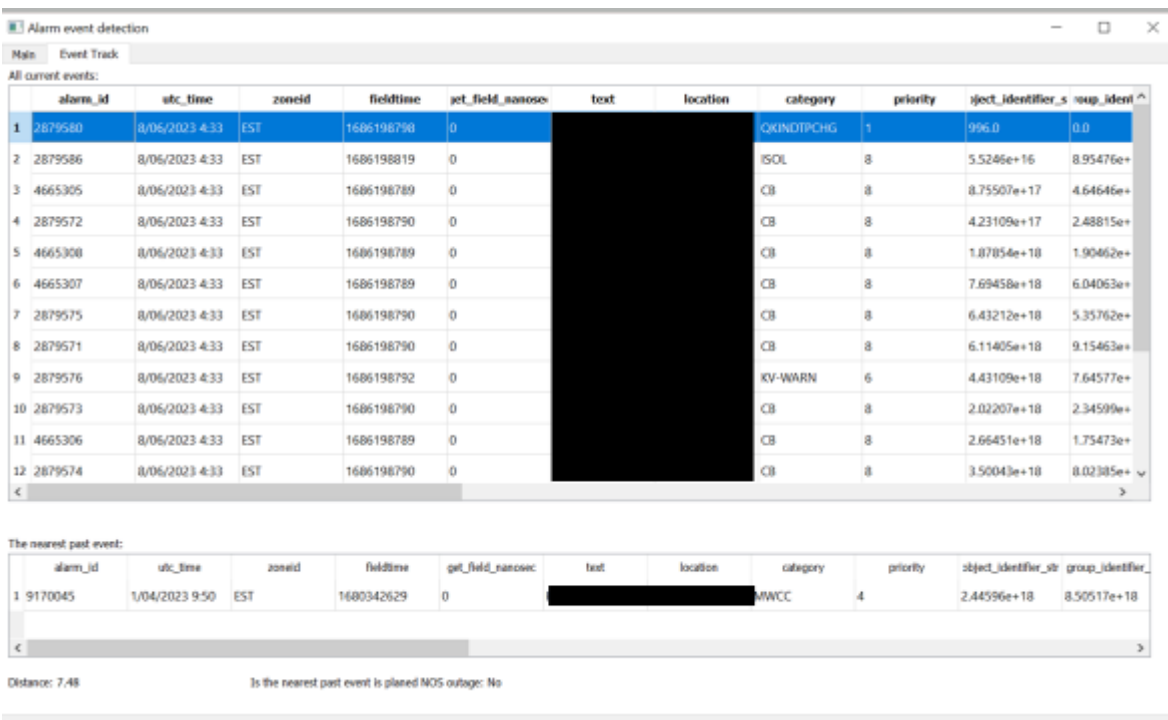
Figure 13 Prototype ML Application for Incident Identification

3.1.2   Selecting and Curating the Operational Datasets

To train the machine learning models on real operational datasets, there is need to use as much operational historical, archive and background data as possible. This process is not automatic and easy. There is a lot of manual data handling work involved in the combination and alignment of the datasets as the databases were in different applications and different data formats. Although time-stamped the actual time reference is not consistent in all the datasets, so data must be time synced manually.

The following datasets and databases and OT applications were used.

EMS/SCADA: This is the primary operational system that receives data from the TNSPs in the NEM via secure communication channels. The information in this database includes switchgear asset

changes, analog numerical data points or limit breach alarms, communication and system data. This data set is the primary one used by operators to monitor the system.

EMS State Estimator: The application in the EMS that logs the network topology and numerical data at regular snapshots in time. This is a text-based output that can be used in other simulation-based applications such as dynamic security assessment applications.

EPSOC Log (Electric Power System Operator Console): A text-based log application where operators log notable events and incidents on the network as they occur with relevant information and meta data.

SMIRK (Systems Market Incident Reporting Kiosk): The application used by operators to log market incidents to events that are published to participants and on the AEMO website.

NOS (Network Outage System): The IT system used to interface and coordinate the planned outage management process information between AEMO and the TNSPs. This has primarily asset information and times of planned outages and is published on the AEMO website.
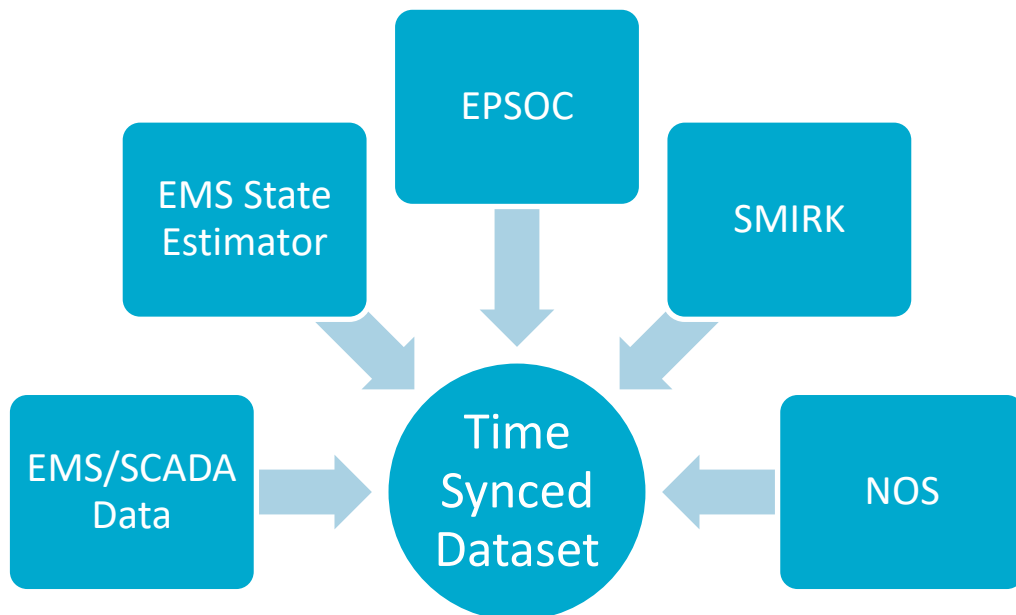
Figure 14 Visual of the combination of the operational datasets

3.1.3 Methodology for Execution of Task 1

Figure 15 shows a high-level process flow for execution of Task 1.
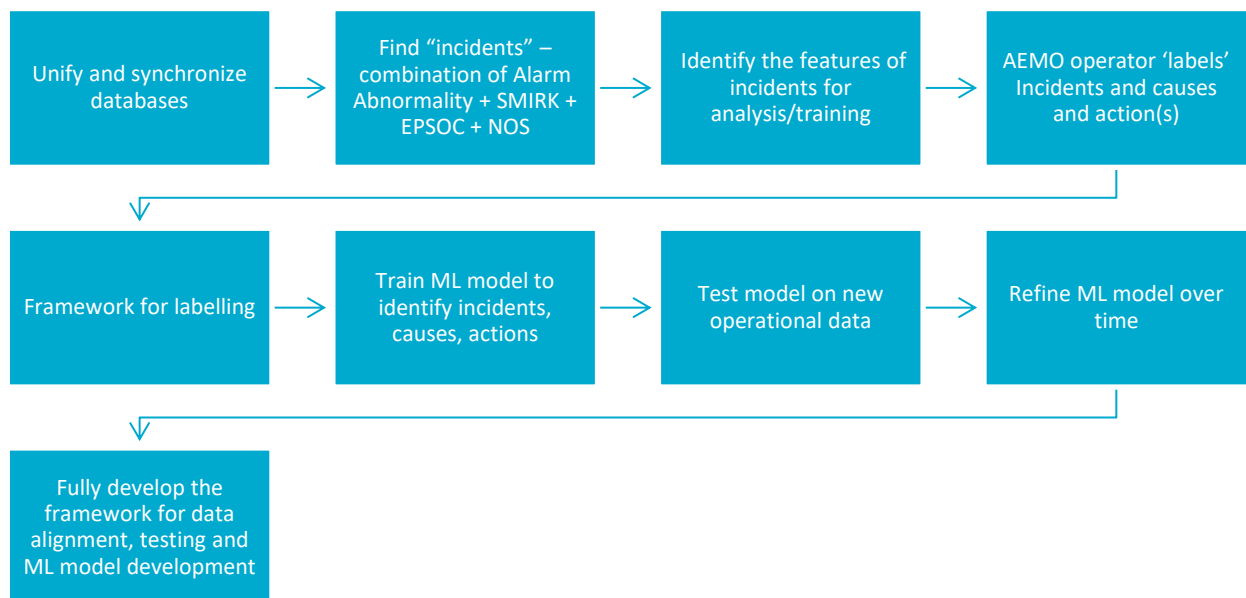
Figure 15 High level process flow for the execution of Task 1.

3.1.4   Combining the Datasets

AEMO have a sandbox environment where datasets can be combined for data exploration.

The alarm data set is the baseline dataset as it is the richest dataset that is dynamic and frequently updated. To begin the trial and process for data curation, a two-month period was selected for use between April and June 2023. **There was a total of 8.8 million alarm records in this period.**

Not all these alarm records are real alarms in the true sense of the word and the majority are alarms from the systems, communications and other non-operational notifications.

The datasets were combined in a commonly used, flexible time synchronised database called **Postgres.** Not all datasets had precise time syncing – the operational data from EMS/SCADA has 1 second granularity while outage data is at 1 minute granularity. Charts and trends were created to visually show the synchronisation of the data as accurately as possible. In some incidents it's possible to visually identify when there is an increase in operational activity.

3.1.5   Pre-Filtering Alarm Data

The alarms that are received in the EMS in the AEMO control center are pre-filtered in the EMS to supress and not show alarm data that is considered low priority or is superfluous to the operator's situational awareness.

The aim of the task is to replicate the operator's direct experience and to develop a tool to build on existing data and tools but that augments their awareness. For that reason, the same filters were applied in the AEMO sandbox, to mimic the actual alarms that appear in the control room for more accurate analysis.

One week of the combined time sync data is shown in Figure 16. One day of data with alarms, NOS outages, EPSOC logs and SMIRK logs is shown, with explanation in Figure 17.
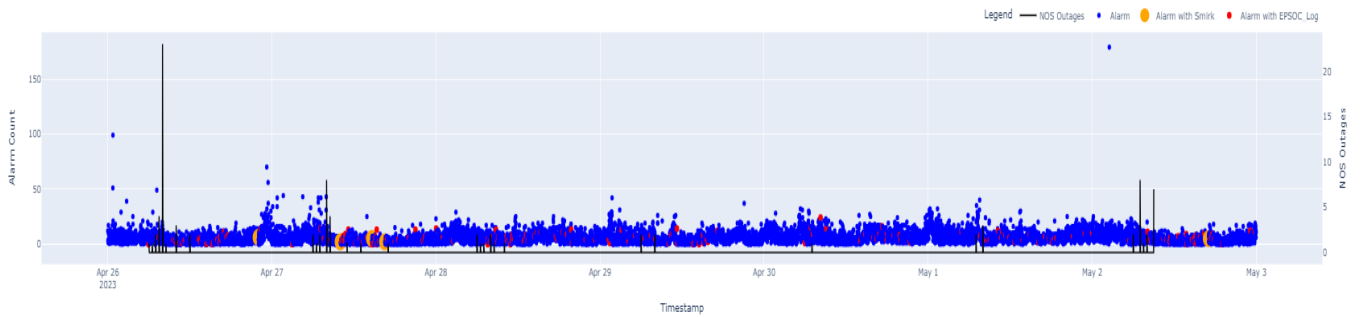
Figure 16 Visual representation of one week of the combination of the operational datasets on a timeline
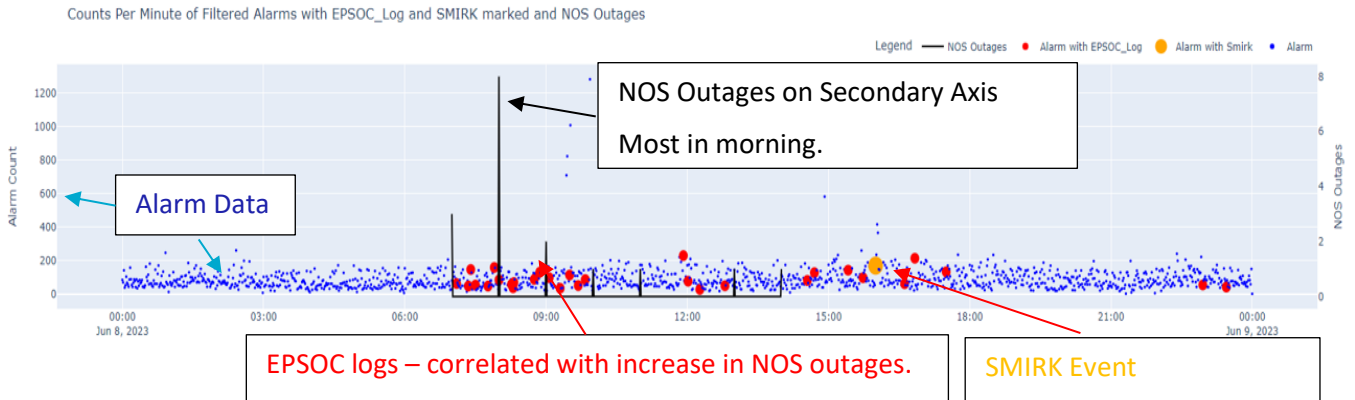
Figure 17 Visual representation of one day of operational data

3.1.6   Labelling the Datasets

Once combined and time synced, it was necessary to add more context to the datasets. The most efficient and accurate way to achieve this is to work with domain experts (in this case AEMO network operators) to eyeball the data and to explain as clearly as possible what is happening when these data points and correlations emerge. The class of machine learning known as supervised machine learning relies on datasets being pre-labelled with an explanation of what the data represents. This allows a model to be trained on this data where it knows inputs and what the output is. Having a dataset that is labelled allows for comparison of ML models when they are tested.

In the prototype for Task 1 - if there is a spike in alarm activity co-incident with a SMIRK event notice, and an EPSOC log entry or NOS entry – a label added to the alarm datapoints would indicate an incident has occurred and additional context.

This is a manual and laborious time-consuming task, but it is a necessary enabler of machine learning in this context. If the application becomes operationally available, it will also be necessary to continuously label and to correct labels as the network changes.

For this reason, it was necessary to develop a framework application for labelling, to allow labels to be added easily, in a structured manner and retrieved and searched. AEMO developed a streamlined process for labelling with their operators and have held several labelling sessions with operators to perform this task. It will obviously not be possible to label all incidents in the entire operational archive within the scope of this project, but the development of the labelling framework will for continued development beyond the scope of the project. EPRI also have an application for the labelling process.

3.1.7   Working Backwards from SMIRK Records to Incidents

Two approaches to labelling with the operator were identified.

a) Work through the alarms or alarm spikes to find incidents.

b) Work from known incidents logged in SMIRK and EPSOC

c) Automatically label with defined windows around known incidents in SMIRK

Option a) was tried but it was found that even major incidents did not, in some cases have large alarm spikes and that important alarms were on-off and may be difficult to identify without in-depth analysis.

Option b) was seen as more efficient for people's time and more representative of actual incidents. With this approach the team identified SMIRK events and found the time stamp for the alarms in the alarm dataset and labelled the important alarms.

Option c) adds more efficiency but is less accurate. By identifying the characteristics of incidents in the SMIRK log, a pre-defined time window in the alarm dataset before and after the SMIRK log time, associated with the stations in the SMIRK can be auto labelled as an incident. This can act as a verification or validation with the expert labeller.

3.1.8   NOS Outage Data

The NOS outage data was parsed into a machine-readable format using the location fields and client composite ID in the EMS database. This outage data is used in the ML application to identify if test incidents were planned or planned based on available NOS data.

3.1.9   Network Model Graph

As part of the research the electrical model of the network was extracted from the EMS state estimator and topology processor. A graph network was created based on geographic distance. This graph network can be extended with more features such as the alarms within the or device level. It is very useful for visualisation of clusters. It was not extensively deployed in this task but will be utilised further in future iterations of the project.
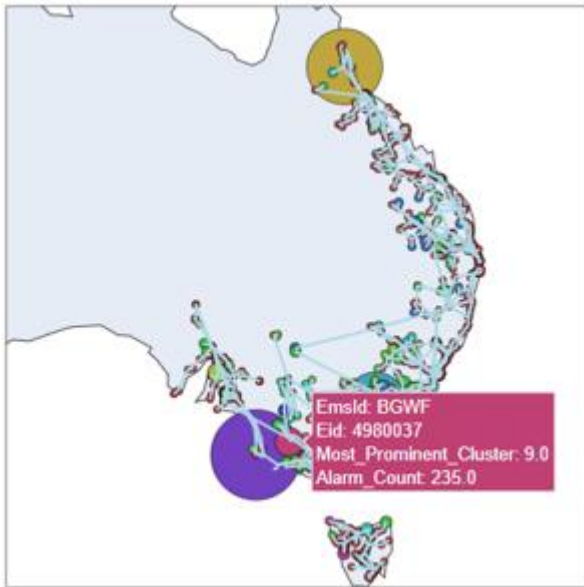
Figure 18 Graph network model based on AEMO alarm data.

3.1.10   Example Incident Labelling Process

An illustrative example of the labelling process is instructive, to explain the process. Two EPSOC logs and a SMIRK log were logged between 15:00 and 16:00 on June 8th for a non-credible contingency, which is an important "incident" meriting further analysis. The event did not see a major uptick in alarm activity, so a statistical analysis would not identify an incident automatically. See data in Figure 19
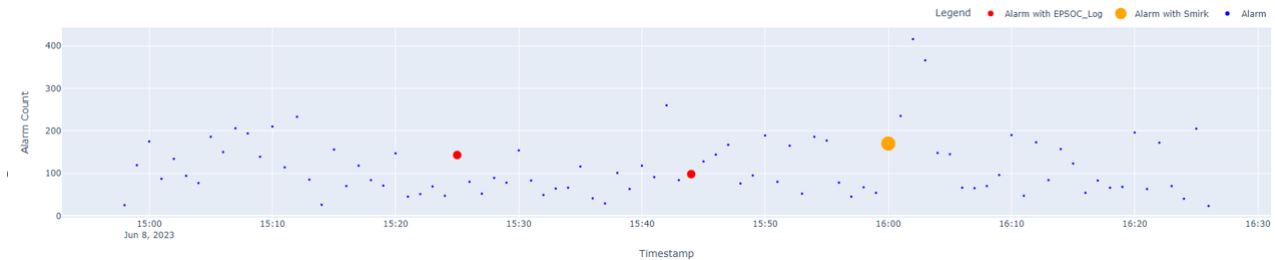
Figure 19 EPSOC and SMIRK logs at 15:00-16:00 on June 8th.

Based on this SMIRK and EPSOC ID the operator could also label the relevant alarm data in the alarm dataset and add a confidence that these alarms represent a true "incident"

Figure 20 Alarm dataset associated with SMIRK event.

An automatic time window of 30 minutes before and after the SMIRK log was also created around with alarms from the relevant stations. This adds further to the manually labelled alarms and can act as a validation check on the labelling. See the rightmost column in Figure 21 which shows two forms of labelling for the same incident, confidence scores can be added to labels, with operator labels given 100% weight while auto labelled alarms given less weight.

Figure 21 Alarm dataset labelled by operator and auto generated by time window.

This process allows for the development of a labelled dataset for use in the ML training.

3.1.11   Machine Learning Model Development

A high-level schematic of the architecture of the ML application is shown in Figure 22. There are offline and online processes.
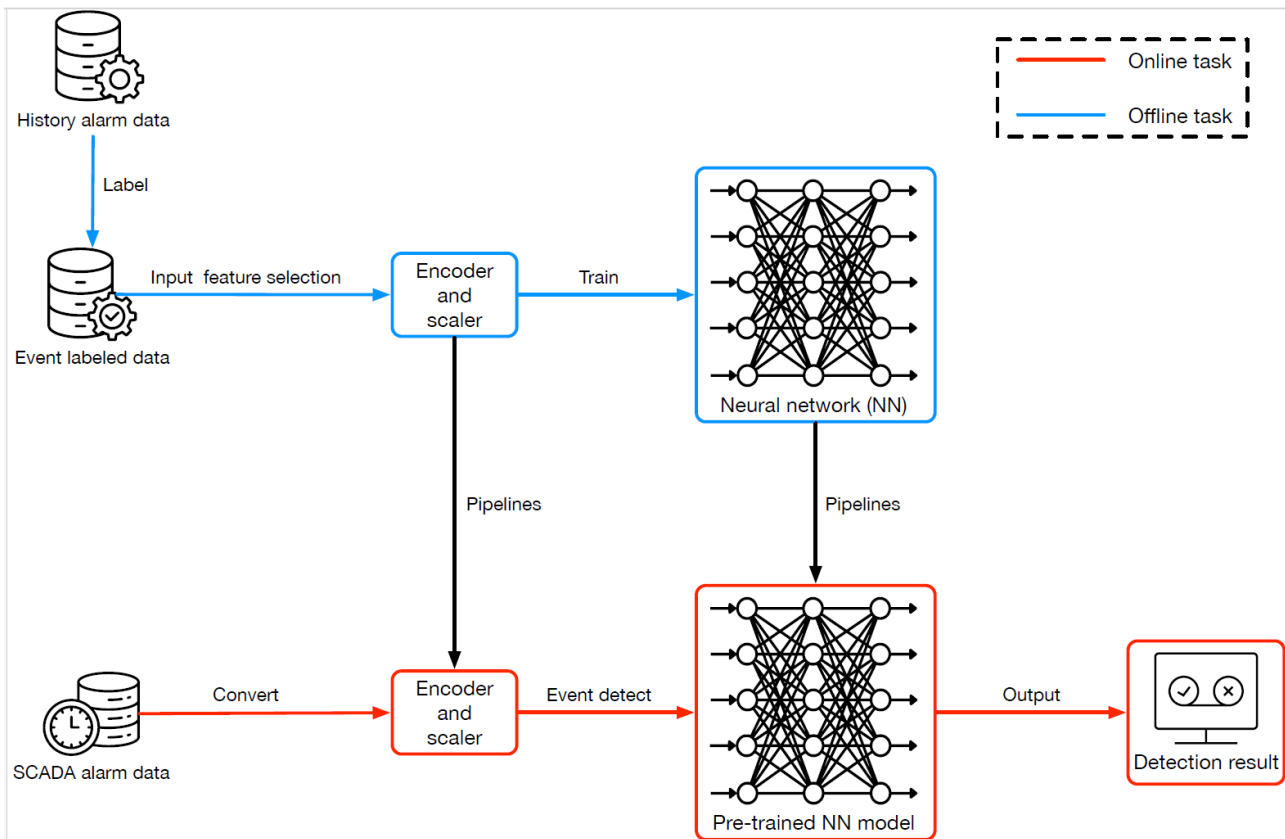
Figure 22 High Level Schematic for the ML application showing offline and online processes.

Training the ML Model

The offline processes involved taking a subset of the data designated for training that is labelled, as described above. These are shown in blue lines in the schematic in Figure 22. A set of features in the alarm dataset are identified for exploration in the ML model to tune the accuracy of the model, these features can be easily changed and added to over time. The initial alarm features chosen are shown in Table 2.

Table 2 Alarm features extracted in model for training.

Alarm Feature	Description
Location	The location of the substation on the network. There is a discrete text string list of hundreds of substations in the network.
Category	The category of the alarm such as breaker operation, communications, generation status, EMS system
Priority	Integer number from 1 to 8 with 1 indicating severity of the alarm when it activates.
Client_Composite_ID	A unique identifier for the asset on the transmission network. Unique text-based string per asset

Exception name	Identifier for the characteristic status of the alarm. Binary status ON/OFF.

These five where features were chosen to be broadly representative but can be changed in future iterations or model improvements.

The characteristic architecture of the ML model used for training was:

- 4 layers: 1 input layer, 2 hidden layer (512, 64 nodes), 1 output layer (1 node).
- Activation function: relu for hidden layers, and sigmod for output layer.

The training/testing split was 97.6% / 2.4%.

For the training dataset the data used was.

- Train data: 3271 records
- Train event: 262 records (i.e., alarms that were marked as being part of an event)
- Train non-event: 3009 records (i.e., alarms that were not marked as being part of an event)

For training of the ML model, the following characteristics were used:

- 100 epochs, batch size is 32.

With this model, the training accuracy was approximately 96% off a small number of data points.
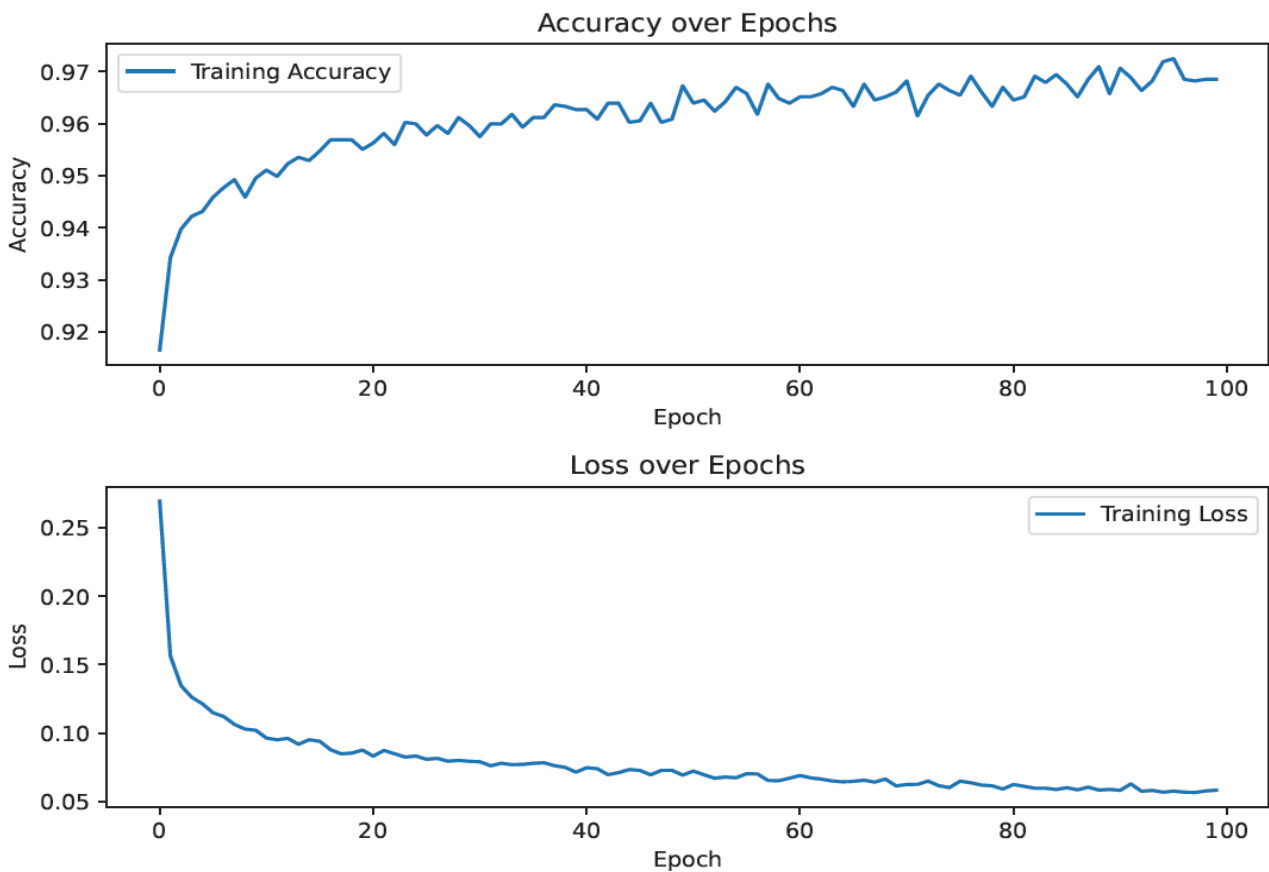
Figure 23 Training Accuracy and Loss over Epochs for the ML model.

Testing the ML Model

For testing of the model, the following data was used:

- Test data: 82 records (2.4%)
- Test event: 16 records
- Test non-event: 66 records

For testing of the model, the accuracy was approximately 93% but the sample size was relatively small (small number of labelled events in 2-month time window). The output of the model was a simple binary yes/no on if the alarm was predicted to be part of an event or not.

The training was carried out securely on AEMO high performance computing systems and was an offline process, meaning the model can be trained and re-trained to improve performance and with more operational data and labels.

Distance Calculation and Metric

AN operator that may use this ML model for real time operations needs to know not just that a real-time alarm can be related to a past archive event – but also – how similar this alarm is to the past event. This requires an estimation of similarity to be adopted. To measure similarity, the distance calculation is necessary to give an estimate for how close a data point is to another datapoint. The text objects are encoded into numeric vectors and distance between the vectors is calculated. In the ML model a **Minkowski Distance** is used with equation shown below.

$$\text{Minkowski Distance} = \left(\sum_{i=1}^{n} \left| x_i - y_i \right| \right)^{1/p}$$

Figure 24 Minkowski distance calculation equation

…where x and y are the vectors and p can be either 1 (representing the Euclidean distance) or 2 (representing the Manhattan distance).

A distance value of 0 represents that the test incident is identical to the trained incident. As the distance value increases there may be some differences in the train and test points. The distance value in the interface should give the operator a steer on how related the observed incident is to past incidents in the training dataset.

NOS Relationship

During testing and experimentation in the task an interesting dilemma was observed relating to outages. If a real incident is detected by the ML model in the test dataset such as an unplanned breaker operation, it does not add much to operator sense-making to link it to a historical **planned** incident. Likewise, if a breaker opens in the test dataset and it is part of a planned NOS scheduled outage, the incident detection ML model should not relate this to a past unplanned incident of the breaker opening.

So, there is a need to build in data related to planned outages so that they can be excluded from the model or at least this information is identified to the operator to be aware of the links. A flag for relationship to the NOS event is incorporated in the model and shown in the GUI described below in Section 3.1.12.

Online Processes versus Offline Processes

The aim of this solution is to be an online application that is continuously trained as it infers and is labelled with events. However, since access to real time data and systems was limited, the prototype was limited to a "one-shot" learning approach (training then deploy) which meant a process for training was separate to the process for testing from a defined alarm dataset. As the research and solution evolves in the coming years, the model for training, testing and validation will also evolve to align with best practices.

3.1.12   Graphical User Interface for the Prototype Testing

A graphical user interface (GUI) was developed for the ML Incident Detection model testing that is described above in Section 3.1.11. The GUI has two main windows for interacting with the ML model described below. The GUI is used for testing (online process) and not used for interaction with the training of the model. I.e., in this prototype application, the model is trained offline and uploaded to the GUI, where it can be tested on data.

Alarm Data Intake Interface

The alarm data intake window is shown in Figure 25. The data intake window allows the ML model that was previously developed and trained to be uploaded and the alarm data for testing to be manually uploaded from a csv file. It is an intuitive easy to use interface.
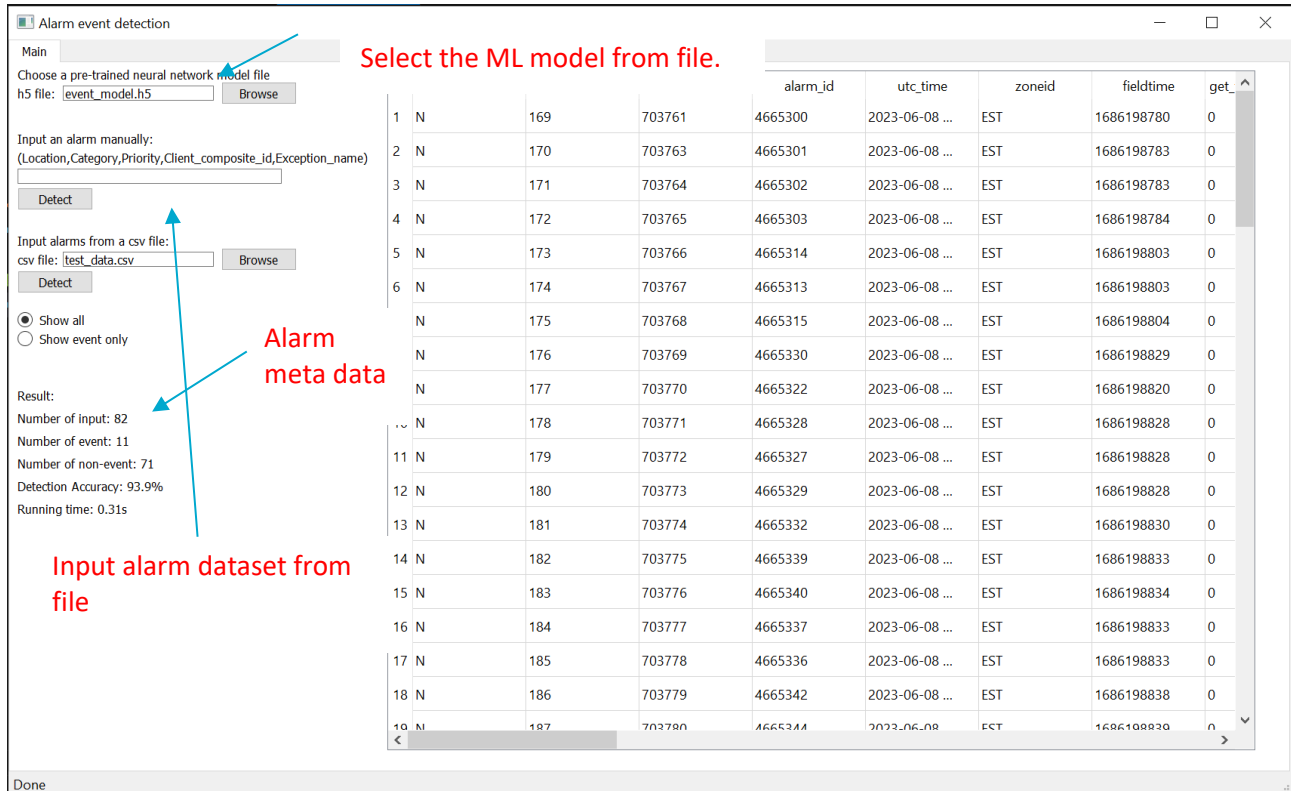
Figure 25 Alarm data intake window

Alarm Incident Analysis Interface

The alarm incident analysis interface is shown and described in Figure 26. It is an intuitive interface that shows the identified alarms related to incidents in test dataset in the top window (where the 12 alarms are shown).

When an incident is selected in the top window the related event that the ML model is most closely related to is shown in the bottom window for reference to the operator. With this design the ML application is identifying **what** has occurred but also **why** it has made the evaluation that an incident has occurred. Usually this will be a reference to a past incident on the network that was labelled, but this is also valuable information for the operator in real time who may need to know when an incident has most recently occurred and other associated alarms.

The value of this approach with ML is that what and why evaluation is almost instantaneous (less than 1 second). This efficiency (reducing the need to search and find past events) will increase the ability of an operator to detect and diagnose incidents when they occur.
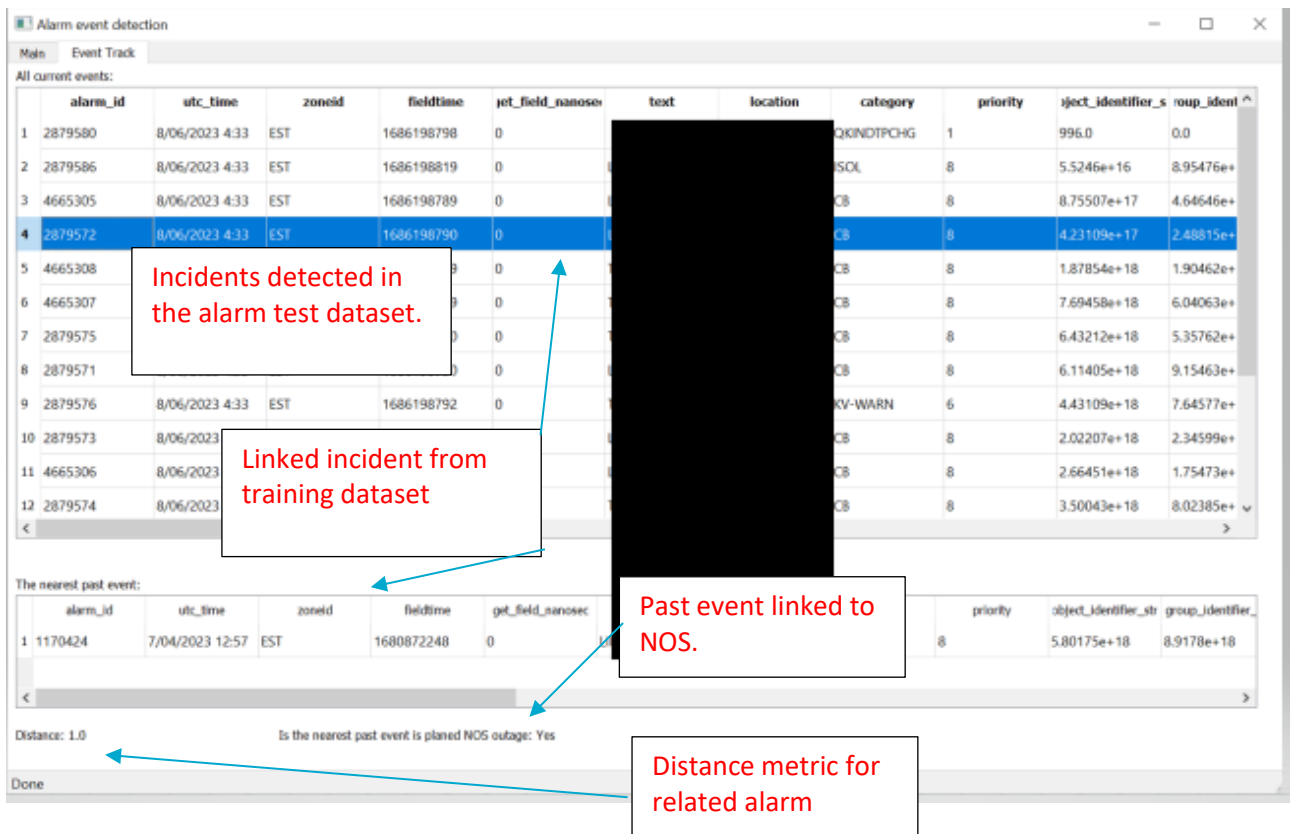
Figure 26 The alarm incident analysis interface

3.1.13   Results

The test dataset was tested and showed a test accuracy of 93% but this was off a small sample size of labelled data. Further tests on more labelled data are required to get an accurate view of the AI model efficacy.

3.1.14   Potential Enhancements

This is a first examination and development of the ML model and shows the power and potential of this approach. However, there are many areas for potential future improvements. Listed below not in order of priority.

- **Model refinement and accuracy:** The model will be tested with new data as it is archived, and the trained model will be enhanced over time with more labelling of events either by an expert or automatically from SMIRK. The five features selected for the ML model may be adjusted to get better accuracy. The baseline test results will improve over time.

- **More complete NOS data integration:** The NOS data is feature rich, time series with electrical model characteristics in-built. NOS data is used in the prototype to distinguish between planned and unplanned incidents, but this can be enhanced further including with integration with the underlying electrical graph model.

- **Linking with the underlying electrical model:** The ML model works off text and time series data. A graph model of the electrical network was also developed, but the features were not integrated with the ML model in the prototype. This enhancement will connect text-based alarm data to the underlying electrical model and connectivity.

- **Developing Suggested Actions:** The initial prototype identifies what has happened, with an estimate of why it has happened. The next iteration should develop a "what now" aspect which would potentially show what actions are needed to resolve the incident. This may be useful for adding constraint equations for contingencies, where there is a record of previously used constraint equations for various incidents that could be instantaneously presented.

- **Deploying LLM:** Text or voice-based queries of the operational data set would be useful to further enhance efficiency and to auto generate reports.

- **Enhanced GUI and visualisations:** The prototype GUI is functional but basic. There is scope to improve the GUI to make it more intuitive and interactive and to include visualisations utilising the geographic graph network as shown in Figure 18.

3.2 Task 2 – Large Language Model Application

3.2.1 Task Outcome

The team created a working prototype of an application to interact with a synthetic alarm dataset via text-based user prompts. The prototype works like commercial LLMs like ChatGPT or CoPilot with accurate results from the prompt. The protype could not be tested on real AEMO data because of security restrictions of deploying open source LLMs with AEMO data but the schema of the synthetic dataset mirrors the AEMO alarm schema and so deployment in future should be straight forward.
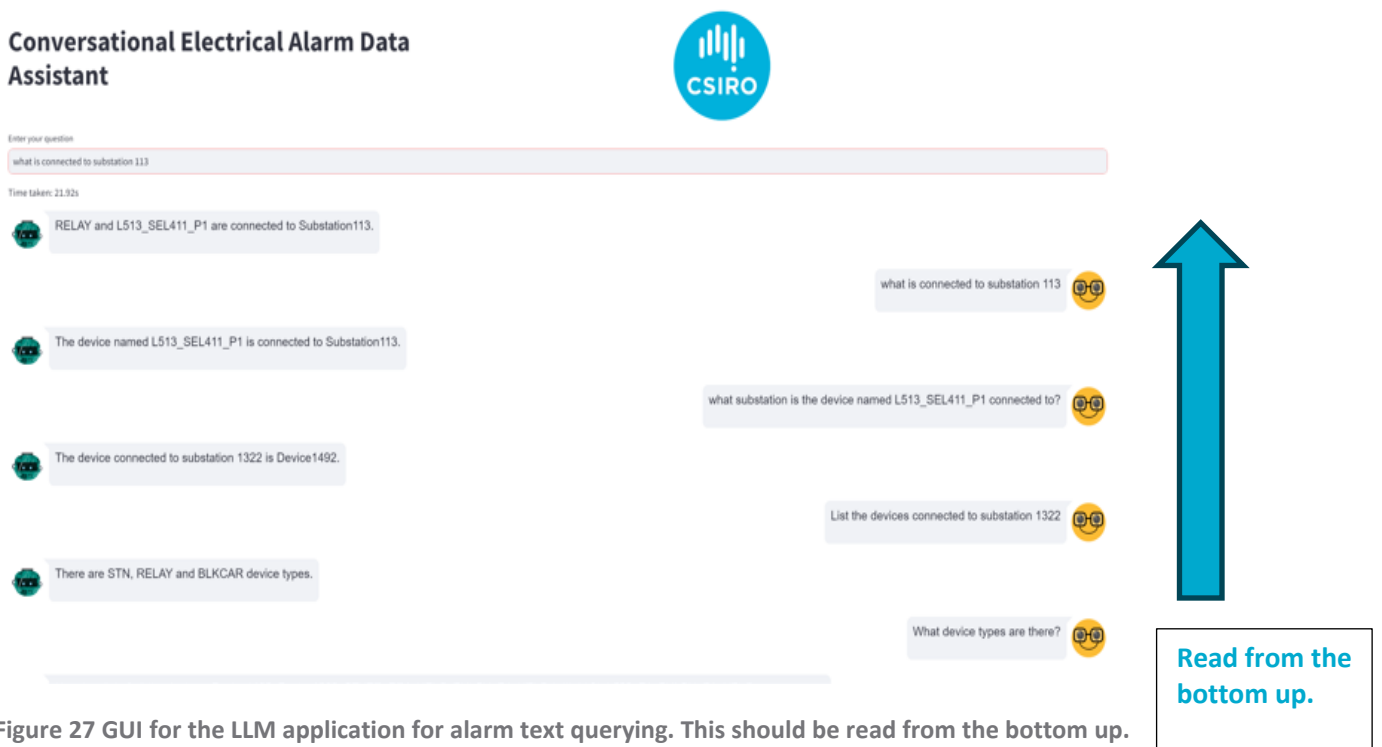
Figure 27 GUI for the LLM application for alarm text querying. This should be read from the bottom up.

Large language models are numerical models of text-based probability weighting of combinations of words appearing together – based on its trained archive dataset[4]. For this reason, it requires a very large corpus of text and language to be trained off. The GPT LLM was trained on the entire text of the internet for example. OpenAI ChatGPT's LLM is not available open source to researchers and its applications are commercial.

Since the release of GPT in late 2022 a number other "openly available" LLMs have been published and made available for research purposes only. Openly available LLMs allow researchers to explore applications of LLMs with a baseline language model, without the need to spend resources to pre-train their own baseline model.

Open source LLMs can also be deployed and trained in offline contexts without the need to be connected to the internet. The project team have explored the use of LLAMA and Mistral LLMs in an offline context and to benchmark performance. To develop applications of LLM in the operational context, in a short time frame, a private LLM is required.

3.2.2   Retrieval Augmented Generation Combined with Large Language Models

While LLMs are trained on a corpus of text data and language, it in general doesn't have additional context for specific applications and contexts that people can use LLM with. Retrieval Augmented Generation (RAG) is a means of connecting a generic LLM (such as the Mistral or LLAMA models) with context specific datasets that are not public (such as operational datasets). Fine tuning is another often mentioned process which means the generic LLM is further trained on domain-specific text data to augment its answers for that specific domain. Fine tuning takes time and costs compute resources. RAG uses just in time connections to datasets using API or database queries to add additional context, so is favourable form the time and cost perspective.

Combining LLM and RAG will provide the basis for testing the efficacy of LLMs in the operational context.

3.2.3   Synthetic Alarm Dataset

Given the emerging and novel nature of LLM technology, and the fact that models are continually trained with prompts and inputted data it was not possible to use real operational data as part of the prototype development. Deploying and using LLMs on AEMO datasets, even for testing purposes is prohibited. A synthetic alarm dataset – the same one used in the Stage 2 CROF project- was used to develop the prototype. A sample of the data and structure is shown in Figure 28. It includes date/time, substation, device, device type, and text-based event message fields. It is readily adaptable to any other alarm schema.

---

[4] Article explaining large language models: Large Language Models | Communications of the ACM

EventId	EventTimeStamp	SCADA_Category	TOC	AOR	Priority_Code	Substation	DeviceType	Device	event_message
1	01/05/2017 00:00	MH	TMS	TOCMSDOC	3	Substation1317	STN	STA_SERVICE	Substation1317 STN STA_SERVICE AC SERV VOLTAGE NORMAL CHG-DET
2	01/05/2017 00:00	ML	TMS	TOCPB	4	Substation1265	BLKCAR	STUTR_PR1_UPLC	Substation1265 BLKCAR STUTR_PR1_UPLC CARRIER LEVEL LOW CHG-DET
3	01/05/2017 00:00	MH	TMS	TOCPB	3	Substation1459	BLKCAR	Device285	Substation1459 BLKCAR Device285 CARRIER LEVEL LOW CHG-DET
4	01/05/2017 00:00	MH	TMS	TOCPB	3	Substation1459	BLKCAR	Device285	Substation1459 BLKCAR Device285 SIGNAL RECEIVED NORMAL CHG-DET
5	01/05/2017 00:00	MH	TMS	TOCGR	3	Substation275	FIBOPT	LABAR_X_311L	Substation275 FIBOPT Device4576 PRI #2 OPTIC/CH NORMAL CHG-DET
6	01/05/2017 00:00	MH	TMS	TOCBM	3	Substation749	RELAY	L37_P2_SEL411L	Substation749 RELAY L37_P2_SEL411L RELAY FAILURE NORMAL CHG-DET
7	01/05/2017 00:00	MH	TMS	TOCBM	3	Substation256	RELAY	L532_SEL421313	Substation256 RELAY L532_SEL421313 SYSTEM NORMAL CHG-DET
8	01/05/2017 00:00	MH	TMS	TOCPA	3	Substation113	RELAY	L513_SEL411_P2	Substation113 RELAY L513_SEL411_P2 RELAY FAILURE NORMAL CHG-DET
9	01/05/2017 00:00	MH	TMS	TOCPA	3	Substation460	RELAY	L461	Substation460 RELAY L461 BKUP SYSTEM FAIL NORMAL CHG-DET
10	01/05/2017 00:00	ML	TMS	TOCED	4	Substation232	BLKCAR	Device689	Substation232 BLKCAR Device689 SIGNAL RECEIVED NORMAL CHG-DET
11	01/05/2017 00:00	MH	TMS	TOCGR	3	Substation1322	RELAY	Device1492	Substation1322 RELAY Device1492 #1 PRIMARY RELAY NORMAL CHG-DET

Figure 28 Sample of the synthetic alarm dataset used in the LLM prototype development.

3.2.4 Methodology for the Development of the Prototype using Graph Networks and LLM

The protype developed uses in-context learning and retrieval augmented generation (RAG) to answer plain language queries of the dataset from the user. LLM is used in three ways:

- The knowledge graph that stores the alarm data was constructed using LLM prompts.

- The user's questions are answered by generating queries for a database.

- The database output is reinterpreted to present it as a plain language answer.

The database is constructed by designing a prompt that leads an LLM to extract relevant information from a database of alarms. An LLM based approach is used in the prototype for this data extraction and graph building but, this task can also be performed efficiently and effectively without using LLMs. LLM was used to build familiarity with structure and code of the graph technology to make querying the knowledge graph more effectively. In future iterations of the research, other approaches will be tested to knowledge graph development.

The type of database constructed is known as a knowledge graph, which is a type of network that represents the relationships between different entities. Knowledge graphs provide a representation of data that is structured to facilitate inference. In this application the entities of the graph are devices and substations and the relationships between the entities are the connections in the electrical network. Neo4j[5] is a widely used graph database management system that is suitable for storing and accessing knowledge graphs. A research license application was used to generate and store the knowledge graph for user queries in this protype application.

Constructing the Knowledge Graph

To construct the database a prompt is engineered that informs the LLM which entities to extract from the data given, and which relationships need to be constructed between these entities. This prompt is run multiple times, once per alarm in the synthetic dataset, and returns the extracted information which is then parsed into cypher commands (the native query and command language for neo4j, the graph database used to hold the knowledge graph).

These cypher commands are used to create the knowledge graph structure in Neo4J. The creation of duplicate entities is prevented by the merge create feature of Neo4j. After all the alarms have

---

[5] https://neo4j.com/

been processed in this manner the result is a knowledge graph structure of the alarm data with clear interdependencies contained in a Neo4j database. This process is shown in Figure 29 (left)

Querying the Knowledge Graph

To allow a user to query this structure - another language prompt is engineered to take the users plain language query and direct the LLM to reinterpret what the user has asked as a knowledge graph query. In-context learning is used to do this, it works by providing examples of what the desired output might look like as well as context about the database schema (naming schemes used etc). Once a knowledge graph query has been generated by the LLM (from the user's input) it is sent to the knowledge graph, the relevant information contained in the knowledge graph is returned to the LLM which reinterprets it in the context of the user's original question and provides a plain language answer. This user query process is shown in Figure 29 (right).
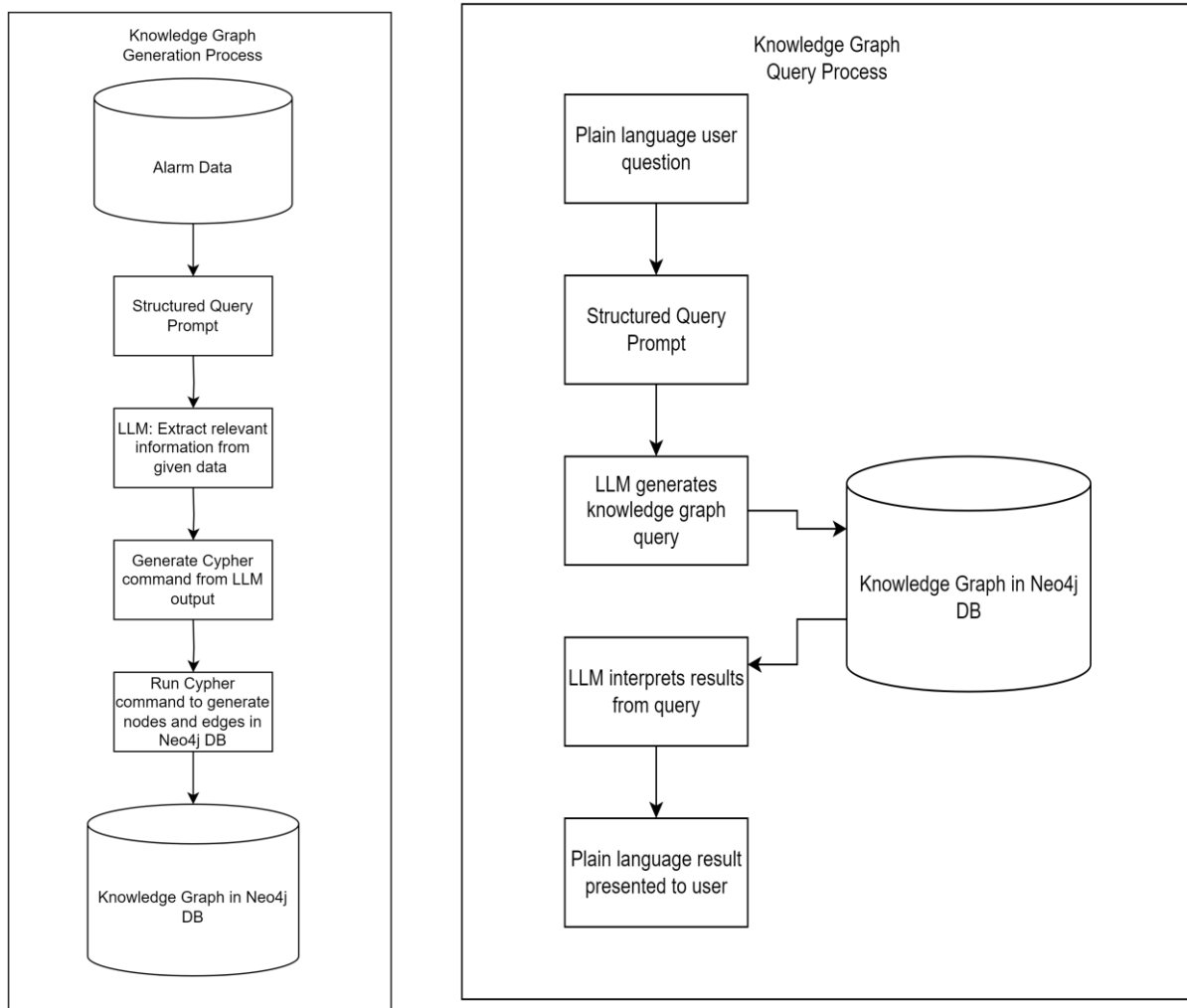
Figure 29 Knowledge graph generation process (left) knowledge graph query process (right)

Deploying the LLM Locally

As described above, there are well established and significant data security concerns with the use of LLMs, especially commercially and security sensitive alarm datasets. To address this issue, it was

decided to run openly available LLMs locally on an external independent machine and link it to the application and knowledge graph database. This means there is no reliance on 3rd party APIs or any need to transfer data on networks not controlled by the user. Running the LLM requires a computer with suitable specifications which can be easily obtained by any network operations company as a locally run machine or a virtual machine from their cloud service provider.

The Llama 3 model from Meta was used in this application for the knowledge graph generation and chat function. The Llama 3 model is publicly available under the meta llama 3 community license agreement. To facilitate the running of the Llama 3 model the ollama toolkit was used. The example application was developed and run on an AWS ec2 instance at RMIT university, the application was accessed through browser on a RMIT laptop.

There was no vector used in the development of this prototype. The open source LLM framework "Langchain" was used to build and connect the different elements of the protype solution in a modular format to make the pipeline more efficient.

3.2.5 Example Prompts and Results

It's important to note again that while the answers to the prompts were broadly accurate, the test protype was artificially set up to match prompts to knowledge graph queries. The accuracy of the of the results is dependent on the ability to infer from the prompt and match to a predesigned query. Typing free form questions into the prompt box would not produce accurate results. Building the prompts and queries and working off larger models will allow for a more robust and continuously improving solution in future.

The prototype was prompted for a list of the devices in the alarm dataset. This was an accurate answer.
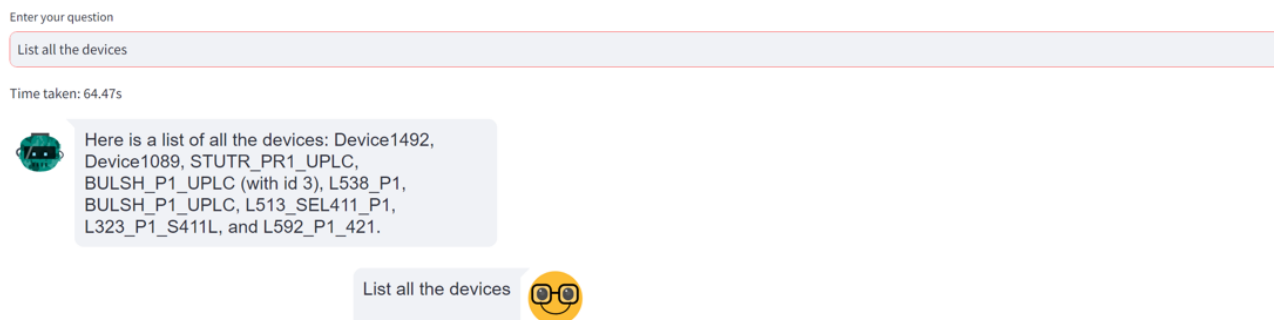
Figure 30 Prompt and answer for a list of devices in the dataset

The prototype was prompted for a list of the substations in the alarm dataset. This was an accurate answer.
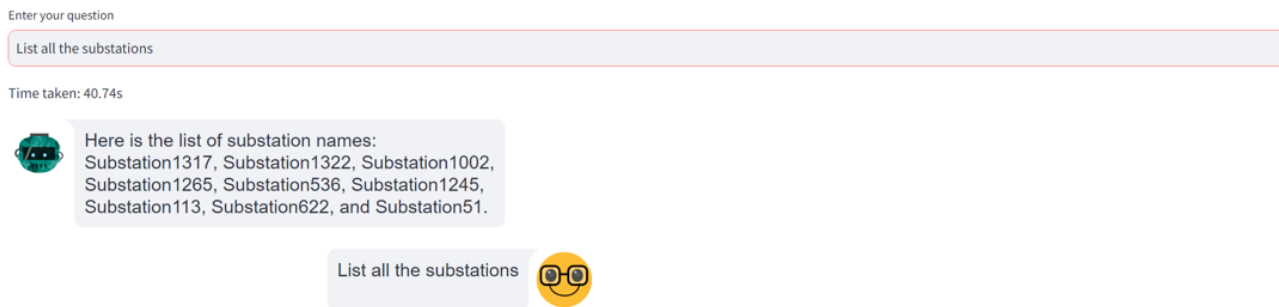
Figure 31 Prompt and answer for a list of substations in the dataset

The prototype was prompted with a specific query about a device in the alarm dataset. This was an accurate answer.
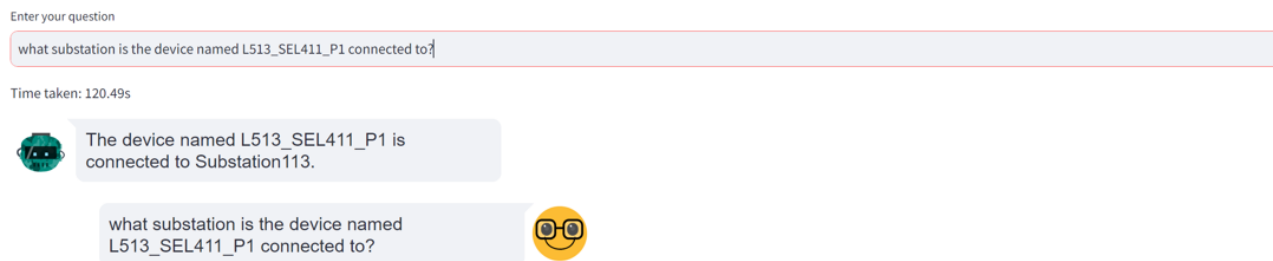
Figure 32 Specific query about a device

The prototype is capable of parsing and accepting multiple prompts and answers in a similar way to commercially available LLMs.
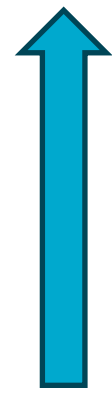
Figure 33 Multiple prompts and answers

3.2.6   Summary and Potential Enhancements

This approach is interesting and potentially very powerful application of LLMs in the operational context, as it shows the possibility of using LLMs to construct informational structures from data, facilitating the use of data to answer user questions. The results seen in this application are encouraging and show the potential for the development of interactive tools that facilitate the interaction between control room operators and electrical system data. Some potential further enhancements,

Prompt Engineering: Currently the approach is sensitive to the prompts used. In the protype only a limited set of prompts that are pre-curated generate real results. Further work is needed to expand the range of acceptable prompts for knowledge graph queries.

Knowledge Graph Generation: There is broad scope for improving the knowledge graph by adding more features from the underlying dataset. In the protype the only the substation and device connectivity is created, but further enhancements would add the time element, priority and other features.

Integrating the Connectivity Model with the Alarm Data: In a similar way to the knowledge graph in Task 1, developing the capability to link a knowledge graph with electrical asset and device connectivity with the alarm data would provide powerful inference for an LLM and add further context to answers. Other data sources, such as text-based logs could also be added.

Multi clause prompts: Developing complex data queries from multi-clause prompts would be useful for the operator in some contexts such as – "When was the last time Device ABC alarms when the voltage on Device DEF was high". Identification of specific use cases related prompts for control room operators could help direct development of this promising approach.

Integrating Task 1 and Task 2 - Automatic Alerts on Real Time Data: Inferences could be made from real time operational data ML models that could be interpreted by the LLM to tell the operator in plain English what the incident is and how it could be solved.

Integrating with Natural Voice Based Applications: Having the ability to use natural language processing to parse voice commands directly into and out of an LLM would improve operator efficiency, especially during periods of high workload.

Fine Tuning LLMs: Emerging research suggests that fine-tuned LLM models may be more accurate than base LLMs+RAG. This methodology and comparison will be explored in future iteration. For synthetic datasets like the synthetic alarm data used, this could be possible but fine tuning an open LLM with a real data set is impossible legally.

3.2.7   Artefacts to be Published with This Report

The source code for the task 1 and task 2 models will be open sourced and published by CSIRO as part of the deliverables for this report. This will enable and encourage other researchers with similar interests to build on the work in tasks 1 and 2 and to improve the models as technology innovations in AI and ML advance in wider industry. The published artefacts will include the commented source code and descriptive methodologies for how the code can be run and deployed.

3.3   Task 3 – Operational Model Validation

3.3.1   Task Outcome

A methodology was developed for the development of an operational digital twin for automated continuous validation and tuning was developed in consultation with AEMO SMEs. AEMO execute elements of the methodology manually on an ad hoc basis, but an automated system is a very difficult and ambitious undertaking. Developing the individual elements of the methodology will be complex and resource intensive but a pathway is established in the report.

3.3.2   Defining a Model Validation Framework

To build on the vision defined for Task 3 in Section 2 a framework/methodology for how the validation process may be achieved is required. The steps in the framework/methodology can be manual as they are now, but it should be possible to automate the individual elements to make the process more efficient. The methodology is shown in Figure 34 and described in the subsequent section.

```
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│ Disturbance     │     │ Disturbance     │     │ Collate dynamic │     │ Collate         │
│ event on        │ ──▶ │ incidence       │ ──▶ │ simulation      │ ──▶ │ operational     │
│ network         │     │ information and │     │ incident and    │     │ data and state  │
│                 │     │ classification  │     │ assess accuracy │     │ estimation      │
└─────────────────┘     └─────────────────┘     └─────────────────┘     └─────────────────┘
         ▼                                                                        │
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│ Validation-     │     │ Collate HSM     │     │ Run dynamic     │     │ Allign HSM data │
│ Candidate       │ ──▶ │ data in         │ ──▶ │ simulation to   │ ──▶ │ to Simulation   │
│ Models and HSM  │     │ vicinity map to │     │ replicate       │     │                 │
│ with relevance  │     │ model           │     │ disturbance     │     │                 │
│ assessment      │     │                 │     │                 │     │                 │
└─────────────────┘     └─────────────────┘     └─────────────────┘     └─────────────────┘
         ▼                                                                        │
┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐     ┌─────────────────┐
│ Define metric   │     │ Determine Model │     │ Create Report   │     │ Engage asset    │
│ and thresholds  │ ──▶ │ Accuracy        │ ──▶ │ on Validation   │ ──▶ │ owner and/or    │
│ for validation  │     │ Compare HSM     │     │ Process and Log │     │ OEM with report │
│ accuracy        │     │ Data to         │     │ Results         │     │                 │
│                 │     │ Simulation Data │     │                 │     │                 │
└─────────────────┘     └─────────────────┘     └─────────────────┘     └─────────────────┘
```

Figure 34 Proposed Framework/Methodology for automated model validation using operational data towards a digital twin.

3.3.3 Description of the Framework/Methodology

Disturbance Incident on the Network

This is the initiating event on the network that would trigger the start of a potential model validation process. This could be a fault on the network caused by an environmental event which causes a short circuit or voltage depression, or a generator trip event which causes a frequency disturbance. It is best if this is a discrete event rather than multiple simultaneous events which may not be as clear to measure.

How this Could be Automated	Monitoring in real time analog operational data (EMS or WAMS) to detect that a notable incident has occurred. Triggers could be set on the analog monitoring. The work in task 1 could also help if the initiating event triggered a text-based SCADA alarm.
Barriers to Automation	Difficulty defining the triggers for the analogues and validating that they are realistic. Getting a spread of locations for triggering events at important nodes but not widespread enough that there are too many triggers
Enablers for Automation	Configurable analog monitoring in EMS or WAMS. Testing and validation on the trigger values.

Disturbance Incident Information and Classification

When the event has been triggered to initiate the validation process, the nature of the disturbance will need to be classified by the system operator, potentially in collaboration with the TNSP. If it was a short circuit fault on the network, the assets impacted will need to be identified, such as location, faulted phases, duration of the fault, distance to fault on the line. If the disturbance event was a generator tripping that triggers an under-frequency event, AEMO will be aware and will know the details. The disturbance incident information will likely need to be saved in the AEMO EPSOC log by the operators in a structured manner for further use later in the process.

How this Could be Automated	Automatically detecting unplanned breaker operations on the network will give the asset information. WAMS data should automatically collect disturbance data. Automatic information sharing from the TNSPs to AEMO for disturbances from HSM (DFR) would help.
Barriers to Automation	Data/information sharing between AEMO and the TNSPs is challenging, in particular HSM data. Getting accurate fault locations on lines if relay information is not available.
Enablers for Automation	Leveraging the research from Task 1 where incidents are automatically identified and classified based on labelled and trained datasets of past incidents.

Collate Dynamic Simulation Pre-incident and Assess Accuracy

It is important to collate the data from the simulation to assess real time simulation application performance. The dynamic security assessment application should have been running in real time at the time of the incident. It runs in the minutes time frame and presents the most critical issues to the operator for action rather than every possible issue with associated data. The DSA will also automatically study worst case contingencies busbar faults and slow breaker clearance faults, but most short circuits are at random distances on lines and not the worst-case incidents.

How this Could be Automated	This should be reasonably automated. DSA studies are automatically run, and results saved in files that are timestamped and should be easy to access. Automatically finding the nearest simulation results based on the time stamp should be straight forward. If no results file exists, it will still be possible to replicate the event from the state estimator.
Barriers to Automation	The time stamp format of the simulation files should be matched with the time stamp of the disturbance incident.
Enablers for Automation	Automation code to identify and gather files based on a time stamp.

Collate Operational Data and State Estimation:

The next step is to collate all the relevant operational data at the time of the disturbance incident. This is mostly EMS/SCADA data and the state estimator simulation results at the time of the incident. The state estimator gives the topology and numerical operational data of the network assets at the time of the event. The SCADA data can be localised to the location of the disturbance rather than pulling every SCADA data point for the disturbance. A defined buffer before and after the event, likely of the order of seconds should be established to limit data size. The size of the time window will be determined based on the disturbance classification.

How this Could be Automated	Like the DSA simulation results, this should be reasonably straight forward to automate as the results of the state estimation and SCADA data should be time stamped and available in an archive that can be searched.
Barriers to Automation	May not be a need for all operational data, generally localised to the location of the incident may be sufficient, but defining the reduced zone may bring inaccuracies. Time stamp format may need to be aligned.
Enablers for Automation	Automation code to identify and gather files based on a time stamp.

Validation-Candidate Models and HSM with relevance assessment

Based on the location of the incident there will likely be several validation candidate dynamic models in the near proximity. Candidate dynamic models may be conventional generation, CLM models or DER_A models, depending on the model validation granularity. It may be necessary to begin this process with conventional model validation, given the size and pre-existing validation information. Depending on how mature the model validation process is, some of the model may have recently been validated and so will not need immediate re-validation. It is necessary to keep a log to determine all the models and how recently they have been validated for comparison purposes. Additionally, validation will not be possible without HSM data in the vicinity, so there needs to be a check on availability of HSM data near to the disturbance incident. If a fault is close to a generation asset it is better for validation. For under/over frequency disturbances the performance of all generators can be validated.

There must be a trigger for assessment whether validation is:

a) Necessary based on how recently models were validated.

b) Possible (based on available HSM data)

Once this assessment is made it can be decided whether to proceed with the process for the incident.

How this Could be Automated	Difficult. Automating this will require the knowledge of how close a model is to a disturbance. This may be based on number of busbar distance or electrical distance. A searchable log of recent validation needs to be

	automatically updated. The location of HSM in the vicinity needs to be correlated with the model location and incident location. The available HSM data will need to be tested to see if it is sufficient.
Barriers to Automation	Setting rules for proximity to incidents for models and HSM Proximity or electrical distance calculation rules Access to multiple systems and combining real network information with location of devices and models.
Enablers for Automation	Complex process may need to be built up through interconnected services or applications.

Collate HSM Data in Vicinity Map to Model:

HSM data is large and HSM data from multiple sources can be difficult to manage, even though it is time synchronised. The HSM data that is available can be pulled into a temporary analysis repository. This can also possibly be achieved in the dedicated WAMS. Not all HSM data is needed, only the HSM data near the models that will be validated will be required. The time stamping should be synchronised on PMU data so it should be straightforward to pull PMU data around the time of the disturbance localised to the area.

How this Could be Automated	Once the location is known this should be straight-forward given the available data in the archive. Time windows either side of the incident will need to be defined but this can evolve based on trial and error
Barriers to Automation	DFR and relay data is more accurate but can be more difficult to collate, synchronise and align with the other HSM data. The freedom to pull relay data may also be limited and more manual as it is owned by the TNSPs
Enablers for Automation	This may be achievable in the dedicated WAMS, and external process and code may not be required, if so.

Run Dynamic Simulation to Replicate Disturbance

In the dynamic simulation application of choice that is used in real time operations, initialise the model with the input state estimation file. This should apply the topology and generator status from the state estimator to the dynamic models on the network. Based on the classification of the disturbance, apply the event in the simulator and collate the results. This can be a short circuit on a line, a generator trip event etc. The information from the disturbance incident classification should be capable of being mapped to the dynamic model and the state estimator. The pre and post event time can be configured to match the pre and post event HSM data window to make it easier to align. The granularity of the samples of the results of the dynamic simulation should match up with the HSM data. The results of the simulation for the model in question will depend on the validation criteria – it may be voltage current or real or reactive power.

How this Could be Automated	Difficult as it requires links and alignment between the state estimator files (from EMS) with the dynamic simulation file and the disturbance classification information.
Barriers to Automation	Difficulty interconnecting different files and different OT applications without APIs or integration layers may need customised links and file transfer with XML which is not best practice design. Difficulty aligning elements of the different models based on naming.
Enablers for Automation	Having a consistent naming convention for assets across all OT and simulation applications will reduce the need to parse data. Having a centralised single source of truth model that is applicable to all simulation packages.

Align HSM data to Simulation Data:

The process should now have two sets of data for an asset model. One set from the HSM (what really happened) and one set from the simulation (what was predicted to have happened). The datasets should be aligned and synchronised so that they can be compared at the same instant, but this may not be possible based on data and sampling frequency. Calculated automations can calculate rate of change between samples to determine when the disturbance occurred. The data from both can be plotted to allow an overseeing engineer to eyeball check that both are aligned.

How this Could be Automated	This should be reasonably straight forward to compare data sets. A calculation based on rate of change of data between samples should indicate when the large deviation for the disturbances took place so that they can be aligned. This may be developed through trial and error. Develop a plot of the results should be straight-forward given the availability of open-source applications and packages for data visualisation.
Barriers to Automation	Ensuring the units of the results are aligned – some may be in thousands and others in whole numbers. If the validation is on real and reactive power a calculation of this may have to be made from HSM (PMU data) to compare with the simulation result.
Enablers for Automation	Data visualisation applications to compare easily and view if there are inaccuracies and for investigation by the engineer.

Define Metrics and Thresholds for Validation Accuracy

At this point it should be possible to determine through visualising it and through calculation how accurate the model is compared to real data from HSM. Metrics for measuring accuracy and thresholds for the metrics should be chosen for the validation process. These can be predefined and applicable for all validation studies or can be bespoke to individual studies. The typical error

metrics like RMS-E or MAPE or rates of change can be used as first effort and refined as the process gets mature. Ideally it would be possible to use multiple validation criteria in the same datasets and to let the user decide.

How this Could be Automated	This should do a delta calculation between the real and simulated data. The user would input their preferred method of validation or chose from a predefined set of validation metrics. They should be able to view the results of the validation and the trends for comparison
Barriers to Automation	Should be reasonably straight forward.
Enablers for Automation	A widespread set of agreed metrics for validation that can be used by the user. Ability to have visualisation and calculated results of the analysis.

Determine Model Accuracy Compare HSM Data to Simulation Data

Based on the metrics and thresholds, the delta between the data sets should be assessed across the time horizon of the simulation rather than point in time comparisons and a metric-based determination of how accurate the model is can be made. The validation metric can be compared with the threshold for model accuracy to determine if further action is required. It's important to state that 100 % accuracy will likely not be achievable – at least in the initial phases of the process, so the threshold for what determines good performance can be established through trial and error.

How this Could be Automated	Simple calculations and final determination of model accuracy based on comparison between real and simulated datasets.
Barriers to Automation	None
Enablers for Automation	None

Create Report on Validation Process and Log Results

On completion of the validation process, the results should be stored for future reference and analysis and an easy to digest, standardised text report with charts and data should be automatically created for review by the SME. A recommendation on whether to proceed to further analysis or to engagement with the asset owners should be made and included in the report. The validation log – referenced in an earlier step should be updated with the results of the validation process, including date, model accuracy and further action and the report.

How this Could be Automated	Creation of text-based reports that collate the data, trends and results data should be reasonably straight forward with automation scripting. LLM could also be leveraged for this process but not required. Templates can also be leveraged that can be auto populated.

Barriers to Automation	Formatting of the report, unless standard templates exist. Maintaining the ability to edit the report by the engineer.
Enablers for Automation	LLMs and text-based automation scripts.

Engage Asset Owner and/or OEM with Report

If engagement with the asset owner is required to amend the model, the report can be issued, and the owner engaged. It is the ultimate responsibility of the asset owner, in conjunction with the OEM to provide accurate models to the system operator so they are required to update the model based on the best information they have. When changes are made to the model the evaluation can be re-run to determine to validate the new model. If this is successful, the revised model should be updated in all simulation platforms within the ISO and the TNSP. If the model validation was successful and performance within limits the registry for model validation can be updated with the date so that validation may not be required for a defined time.

How this Could be Automated	Automated report and email generation that can be edited by the engineer doing the validating.
Barriers to Automation	None
Enablers for Automation	None

3.3.4   Challenges with the Framework

There are significant challenges with achieving the automated version of the framework. These are mostly IT process automation related challenges such as data sharing, application development and testing, data sharing between network entities and AEMO. However, a version of this process is carried out manually and, on an ad-hoc basis for model validation everywhere. The difficulty stems from the automated aspects, but most individual steps are technically not challenging individually. The enablers and barriers to automating individual elements of the process are documented above.

3.3.5   Composite Load Model Validation

It may be possible to apply the framework to the composite load model and aggregated DER models, but this also has several challenges.

AEMO does not model the distribution network or parts of the sub-transmission network in some places. This means that if models do not exist of the underlying network the disturbance incident cannot be replicated in the simulation comparison exercises.

While it may be easier to change the CLM and DER-A models as they are under the control of the system operator. The challenge is in availability of HSM at the feeder heads for validation. Since PMU devices generally do not stretch to the Transmission and Distribution interface or even on transmission assets.

3.3.6 Future Vision for Automated Model Validation with Suggestions

This is an exciting area for innovation and research and development and no out-of-the-box vendor solutions exist for this process automation. There is a pathway for the full automation of the process including the suggestion of actions for a refined model, potentially using machine learning or other algorithms is the final step. This final step would help asset owners and OEMs to

| Framework and Process Definition | → | Development of Process to Support Manual Validation | → | Automation of Process | → | Scaling to incorpote more HSM and simualtion capability | → | Suggested model parameter changes based on simualtion |

Figure 35 High level pathway for automated model validation

3.3.7 Testing the Model Methodology

While every effort was made to try to test the framework and methodology with existing models, it was not possible as part of this project stage due to data sharing limitations and difficulties working on the model process from an external perspective.

During discussion AEMO engaged with the concept of real time operational model validation and considered it a worthwhile goal. However, they emphasised the major challenges associated with achieving even partial automation of the sub-processes and stressed the need for a long-term approach with appropriate resources. The framework methodology gives a good starting point for developments in this space and to engage with stakeholders on improvements to the process.

.

3.3.8   Potential Enhancements for the Framework for Model Validation

The project team will remain engaged with AEMO SMEs on this concept and with the team of researchers working on the Topic 2 Dynamic Stability Tools roadmap as well as other industry stakeholders and vendors in the modelling domain with a view to advancing the vision and the methodology and testing the efficacy.

4 Research Relevance to Australia

Machine learning and AI are not intended to be considered as a panacea for all problems in the operational domain. They should be a tool or application that can be used to help in multiple operational processes. They should be considered to an end rather than an end in and of itself, in a similar way to how classical engineering solutions such as optimisation are solutions to engineering problems.

They are best deployed as part of a service to an operational process rather than as an autonomous system controlling all processes. AI/ML has had very limited adoption in operational and control contexts globally, mostly because of trust issues on high reliability systems and lack of maturity in the applications and need for compute power.

AI/ML are efficient in the sense making mode of operations - at identifying correlations and patterns in archive data, in data that operators may not have the ability to identify. In sense making AI can help with problem detection and root cause analysis based on past event training,

AI/ML are also very useful in the action making mode of operation. Especially with recent GPT and LLM innovations, they also have strong capabilities for text generation based on archive text datasets so are ideal for reporting, switching plan generation and dispatching for field crews.

When engaging with control operators the key themes that regularly come up where improvement is required are:

- Alleviation of data overload, alarm rationalisation or reduction and combining data from different systems
- Overly manual administrative processes such as data entry that take time away from decision making and awareness.

As mentioned, AI/ML is efficient in these domains, so it is relevant topic for exploration for all operational environments of high reliability organisations.
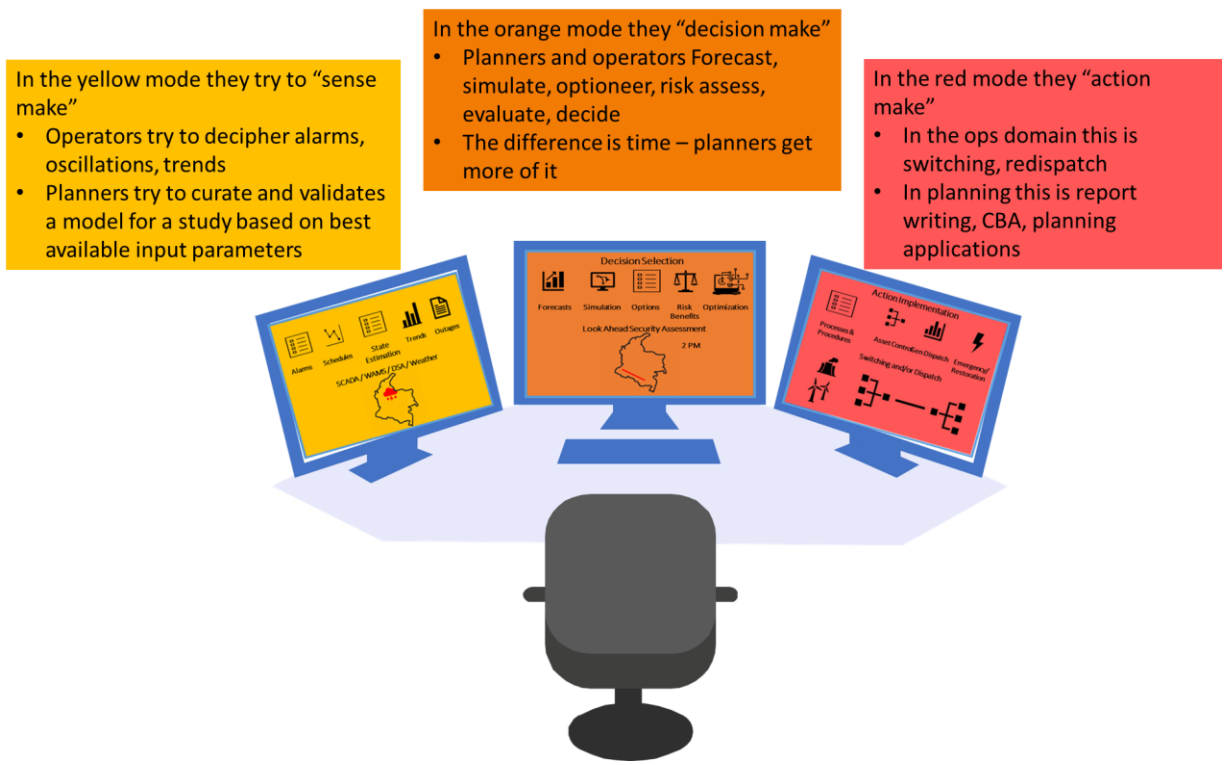
In the yellow mode they try to "sense make"
- Operators try to decipher alarms, oscillations, trends
- Planners try to curate and validates a model for a study based on best available input parameters

In the orange mode they "decision make"
- Planners and operators Forecast, simulate, optioneer, risk assess, evaluate, decide
- The difference is time – planners get more of it

In the red mode they "action make"
- In the ops domain this is switching, redispatch
- In planning this is report writing, CBA, planning applications

Figure 36 A schematic overview of the three cognitive processes for operators

AI/ML Can Help in all Three Modes, but in different ways and with different classified approaches

Supervised learning can help sense make – identify anomalies, validate, check quality and to curate models based on large archive of data. LLMs can also be useful here with text based data

Reinforcement learning can help decision making, trained offline with a simulator on large archive of datasets with decisions/actions

LLM/NLP applications can create reports, plans, dispatch, and suggested actions

Figure 37 The classes of AI/ML that can support the operator int he controls room in their three modes of operation.

In today's world and with an eye to the future decarbonised transmission networks, the power system has essentially digitised.

- Large electromechanical resources have been replaced by smaller power electronically controlled inverter-based resources.

- Large transmission assets and infrastructure are being augmented (but not replaced) by grid enhancing technologies such as special protection schemes, dynamic line rating, FACTS, HVDC

- Electromechanical demand has been replaced by inverter interfaced demand with smart meters tracking energy usage in minute detail.

The available data from markets, resources, demand users, assets and power electronic and inverter interfaced devices and protection and control devices has exploded higher and will continue to explode higher in Australia.

Increases in data, with constant human operators requires smarter IT/OT applications to parse the data and to help the operator make sense of the and gain insights. AI/ML technology is very useful in contexts with large datasets, regular patterns and structured processes.

This means that AI/ML applications will help operators, especially in the sense making and action making modes.

Currently there is no existing machine learning project methodology for power system use cases. In addition, the despite widespread industry and societal adoption, there are limited machine learning applications in the power system and energy sector more broadly, both in Australia and around the world. The development of the methodology and use cases in this project can be used by researchers and practitioners in Australia. The solutions developed in this project for incident detection can, in theory, generalise to other network operators in Australia and globally with similar sets of databases (reporting, alarms, outages etc).

The model validation task is also highly relevant to other network operators in Australia and beyond. As network become weaker, new phenomena arise on the network that need to be studied, this requires more advanced simulation capability which requires more advanced and accurate models. Having a methodology to validate models automatically and continuously with real time data.

5 Progress Related to Research Roadmap

Some early-stage progress has been made on two of the original 2021 research roadmaps for data. The AI/ML techniques were envisaged as a later stage of development, but early-stage work – such as the work in the stage 3 project - must be carried out over many years to achieve success in this field. Elements of the data roadmap that have been progressed are highlighted in yellow in Figure 38 below. It should be stressed that AI/ML applications are long gestation, resource intensive applications, especially in emerging fields such as LLM and especially where data security is of paramount importance.

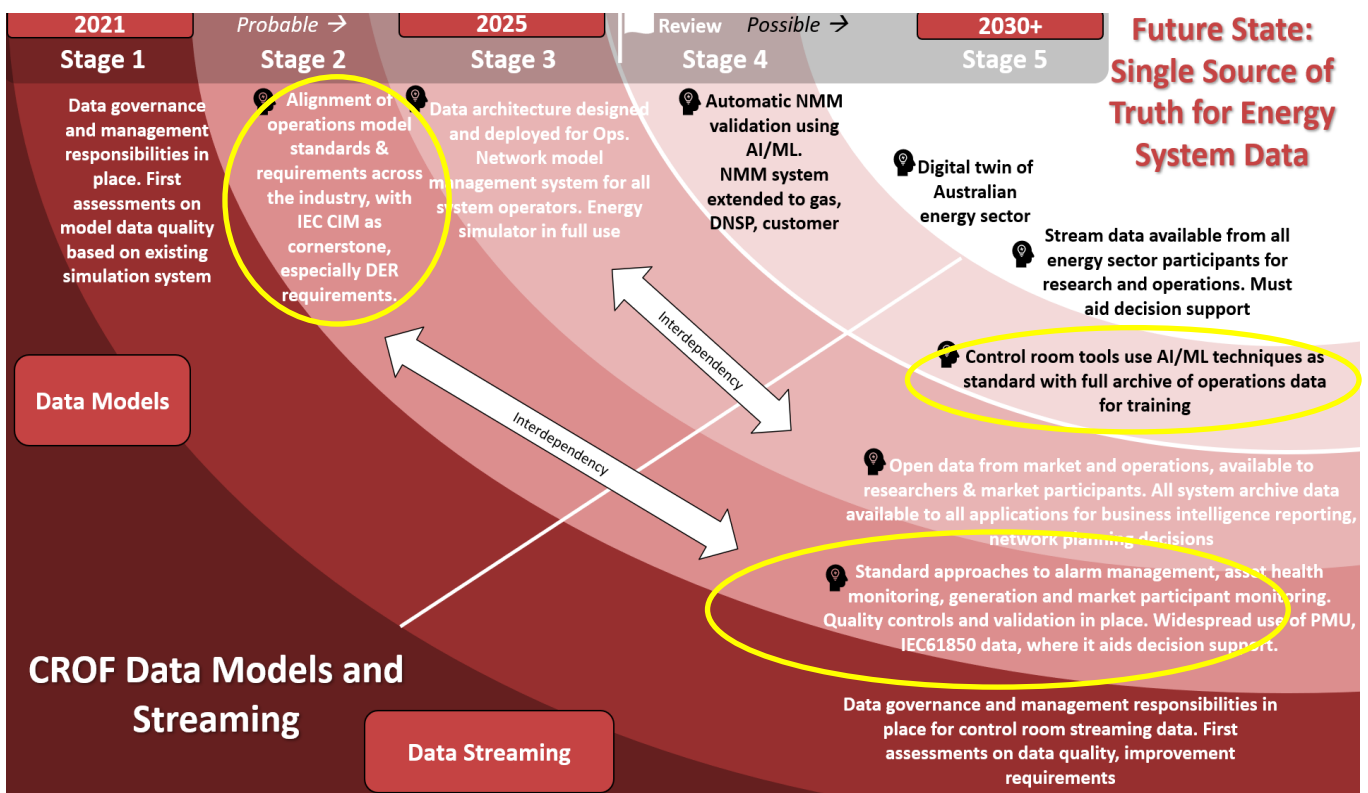Some work around data model standardisation must still be completed in the coming years.

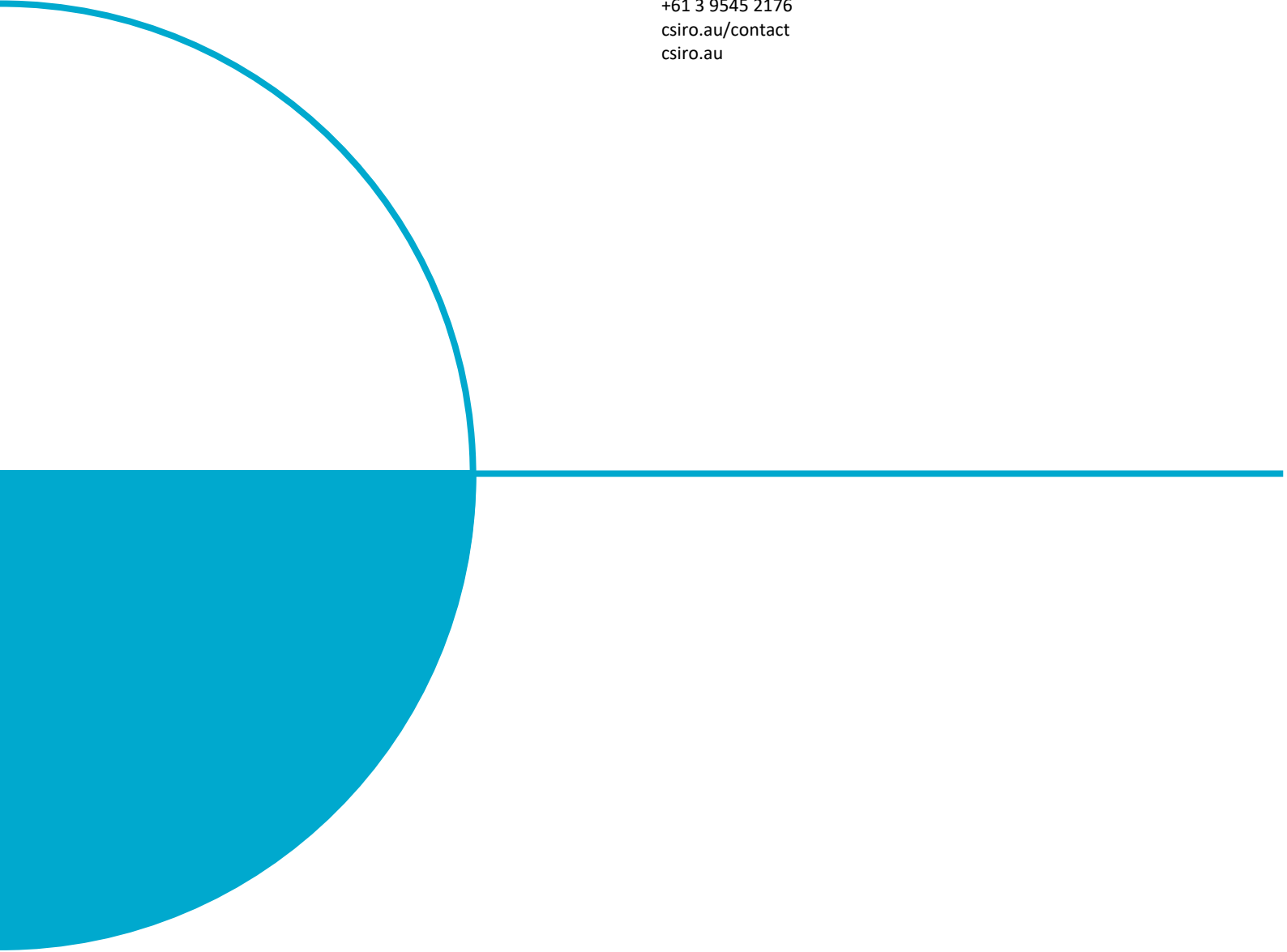**Figure 38 CROF Data Models and Streaming Roadmap**

6 Recommendations for Next Steps and Future Research

The aim of the stage 3 project is to continue development of the incident detection AI/ML model to a fully workable prototype in operations at AEMO and available to other network operators in Australia. Work in this should continue beyond the scope of the stage 3 project but perhaps focussed within the network operators as they transition to business-as-usual operational technology systems.

The 2021 research roadmap was ambitious and vast, covering six core pillars. The focus of stage 2 and stage 3 was on the data pillar, but research activities should also be initiated in the other pillars – Architecture, EMS/SCADA, Operational Technology Tools, Human Factors, Buildings and Facilities.

For future research areas in Topic 3 CROF, it may be appropriate to initiate the human factors research actions, which are focussed on decision making, training standardisation, visualisation and developing the capabilities of the future operator. Additional potential research areas to be explored beyond this project are on the facilities and equipment pillar which focusses on the value of ergonomics and building design to the control room experience and the need for operational readiness centers.

The architecture pillars and the software applications tools roadmaps are less urgent as priorities. They both have active work actions within the CSIRO G-PST framework (Topic 7) and within AEMO through the Operational Technology Roadmap and Programme, co-funded by CSIRO.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.