1 Description of the use case

1.1 Name of the use case

ID	Application Domain(s)	Name of Use Case
UC2.Power Grid	Power grid	Sim2Real, transfer AI-assistant from simulation to real-world operation

1.2 Version management

Version Management			
Version No.	**Date**	**Name of Author(s)**	**Changes**
0.1	29.01.2024	Bruno Lemetayer (RTE)	Initial document (copy from last version of short template document)

1.3 Scope and objectives of use case

Scope and Objectives of Use Case	
Scope	Power grid real-time operation and operational planning (hours-ahead). It integrates in the global concept of assistant framework (developed in UC1.Power Grid) and deepens a specific “real world” complication (in comparison, UC1.Power Grid has a more “theoretical” vision).
Objective(s)	This use case is to assess the capability of an AI assistant to be used for the operation of a “real” transmission grid, in the sense that the “real” environment doesn’t exactly behave as the one available to the agent (that is implemented in the AI assistant) during training and simulation procedures, even if they share the same functional properties (same grid components and topology), and operational constraints. Therefore, Sim2Real stands for “Simulation to Reality”. The main objectives are: 1. Look at additional technical considerations to succeed at deploying an AI assistant in the real world, besides its sole ability to find solutions to simulated situations. 2. Improving human trust when such systems are getting deployed in real-world environment. 3. Allowing for iterative human-AI refinements with human feedbacks and insights.

1.4 Narrative of use case

Narrative of Use Case
Short description
The use case outlines two paths for an AI assistant to manage a transmission grid: A) In coping with real-world conditions, the AI assistant monitors grid situations, raises alerts for human intervention, and provides action recommendations, considering uncertainty from noisy and partially missing data. The human operator makes decisions based on AI suggestions, with feedback loops to continuously improve interactions and learn from realized actions. B) When data limitations prevent full autonomy, the AI assistant alerts the human operator due to missing or poor-quality data. The human operator may also choose actions that do not yield expected results due to various factors. In such cases, the operator can provide missing information to aid the AI. Enriched context, including human input and decisions, is logged for continuous learning, enhancing the AI assistant’s robustness in making recommendations for grid actions. **This use cases only addresses congestion issues**, even if other types of issues can arise on the Transmission Grid and are handled by the operators (e.g. voltage). *Note 1: This use case is linked with the broader notion of “transfer learning”, which is the possibility to adapt a pre-trained model to a new environment only with a slight additional training. One of the possible associated research questions is to evaluate the minimum amount of real data that would be needed in order to align a model with the “real world”.*

Note 2: As for the AI-assistant training, the human operator's decision and perception will rely on "theoretical simulations" (training and simulation tools).

Complete description

The use case can be divided into two paths:

A. The AI assistant copes with real-world conditions

The AI assistant is still able to carry out its role and provide the human operator with action recommendations, even if data is noisy or partially missing.

1. The AI-assistant monitors the situation of the transmission grid *[same as in UC1.Power Grid]*

2. When anticipating issues requiring intervention, the AI-assistant raises alerts for decisions at the appropriate horizon (e.g. few hours ahead to 30min ahead) to the human operator, in time for carrying out corresponding actions *[same as in UC1.Power Grid]*
The action recommendations from the AI assistant will reflect the additional uncertainty due to noisy and partially missing data, and sensitivity to uncertainty.

3. For a given alert, the human operator receives action recommendations from the AI-assistant, with information on predicted effect, and reasons for the decision *[same as in UC1.Power Grid]*

4. The human operator chooses a proposed recommendation, or requests new information or explanations, or looks for a different action guided by an exploration agent or via manual simulation using other specific tools (that aren't part of the AI assistant) *[same as in UC1.Power Grid]*

5. The human operator performs needed actions according to his/her decision *[same as in UC1.Power Grid]*

6. The decisions made are logged with their corresponding context to continuously learn from realized actions and improve the interactions between human operator and the AI-assistant (e.g. relevance of proposed recommendations for actions) *[same as in UC1.Power Grid]*

B. Real-world conditions require specific interactions between AI assistant and human operator

Available data doesn't allow the AI assistant to provide the human operator with action recommendations in a fully autonomous way and requires the AI assistant to call for additional feedback or information from the human operator : the AI assistant raises an inaccuracy alarm.

1. **First type of situation** is where the AI-assistant can't evaluate the need for action due to missing and bad quality data and thus can't determine any action recommendations. It raises a corresponding alert to the human operator.
The main reasons can be:
   - Missing or noisy data:
     - Due to uncertainty because the forecasts aren't always accurate or even available, or uncertainty as "epistemic uncertainty", which is the model uncertainty due to sampling (or underrepresentation) problems
     - The state estimator does not directly use the measurement values but first goes through a readjustment. This means that the raw measurement values from the Energy Management System (EMS) can't be directly used to compute the load flow because the needed adaptations (missing or wrong measurement values due to, e.g., measurement device issues) performed by the state estimator will be missing.
   - Evolution of the electric system: trends such as higher renewable penetration or consumer behavior change (adaptation) that shift data distribution over the years.

2. **Second type of situation** is where the human operator chooses a recommended action that doesn't have the expected consequences on the transmission grid's state.
Main reasons can be:
   - Reproducibility of remedial actions, one or several prerequisites needed to perform an action recommended by the AI-assistant are missing, due to:

        - Device failure (e.g. the failure of a circuit breaker might prevent changing the topology as proposed).
        - Unavailability of flexibility (that might prevent performing planned redispatching).
    - Real-time behavior of the transmission grid is significantly different from simulation due to:
        - Different load flow calculation than the one available at training and inference time.
        - addition or upgrade of new elements on the grid: substations, lines, etc., even automatic devices.
        - Distributed energy resources (DER) can have impacts on grid congestion as well as on decision-making, since they can be a source of additional complexity and difficulty: a model might not be able to analyze or predict the real-world cumulative effect of smaller grid-connected assets.
        - changing grid equipment characteristics (e.g. climate impact, or DLR).
        - transient grid dynamics that steady state simulation don't capture, for example in the context of a windstorm.
        - cyber-physical considerations with the integration and modelling of more automatic devices.

3. When the AI-assistant can't evaluate the need for action, or a recommended action hasn't the expected consequences, the human operator has the possibility of providing the AI assistant with specific missing information to help the AI assistant forecasting system state and assessing action recommendations.
   This is only possible if missing information can be easily provided by the human operator to the AI assistant (i.e., it doesn't generate an important additional workload), e.g., status (open/closed) of a given busbar coupler.

4. The difference between original context used by the AI-assistant and the enriched context are logged to continuously learn from realized actions and improve the robustness and novelty of recommendations for actions by AI assistant.
   Enriched context includes at least:
    - information given by the human operator.

decisions taken by the human operator (visible as topology changes or other actions on the transmission grid).

Stakeholders

See UC1.Power Grid

Stakeholders' assets, values

See UC1.Power Grid

System's threats and vulnerabilities

Human manipulation: Human operators with malicious intent may attempt to manipulate the AI system by providing misleading feedback or deliberately misusing the AI learning process. It is important to ensure that this co-learning process complies with regulatory requirements and industry standards for power grid management.

Adversarial data attacks: Malicious actors might attempt to manipulate the AI system by introducing misleading data or injecting false information into the recommendation process, e.g., feeding deceptive information about the state of a particular grid node, causing it to recommend inefficient solutions or worsening congestion; or, injection of a sequence of false information to flood the human with requests during peak grid operation times.
Trust from human operators: The operational performance of the AI assistant will not be close to 100% of problems solved, which may hinder the confidence and trust of the human operator in the AI recommendations. This will introduce a negative cognitive bias in humans.

1.5 Key performance indicators (KPI)

Name	***Description***	***Reference to the mentioned use case objectives***
Robustness to uncertainty	Describes the ability of the AI system to maintain its performance under varying conditions and in the presence of uncertainties and noise that were not specifically accounted for during training. Evaluated by testing the policy in different scenarios, not represented in	Objectives: 1,2,3

	the training data, by computing the KPIs from UC1.Power Grid.	
Transferability across fidelity levels	Measures how effectively a policy or model trained in one environment (low-fidelity simulation) performs when applied to different environments (e.g. high-fidelity simulation or real-world operation). Evaluated by directly applying the policy trained in a low fidelity simulation to high fidelity simulation and measuring its effectiveness by computing the KPIs from UC1.Power Grid.	Objectives: 1,2,3
Generalization to different grids	The ability of a policy to perform well in an unseen grid environment, which was not part of the training experience. Tested by exposing the previously trained AI system to different environments with changed grid elements and observing how well it adapts and performs by determining the KPIs from UC1.Power Grid.	Objectives: 1,2,3
Assistant relevance	It is based on a comparative evaluation by the human operator of the scenario to be run in the following cases • without AI assistant • with AI assistant Each scenario will get a grade ranging in [0, 10] for each of the 2 cases, given by the human operator at the end of the scenario, with: • 0 corresponding to the worse possible operating condition from the operator's point of view, • 0 to the best possible operating condition from the operator's point of view.	Objectives: 2
Assistant disturbance	It aims at measuring if the notifications raised by the AI assistant are disturbing the activity of the human operator. For each notification, the score ranges in [0, 5] with: • 0 meaning that the notification was not considered disturbing at all by the human operator, 5 meaning that the notification was considered as fully disturbing by the human operator.	Objectives: 3
Workload	It is based on a workload assessment of the AI assistant by the human operators. It shall be determined according to the NASA-TLX[1] methodology or similar[2].	Objectives: 3
Assistant alert accuracy	It is based on the number of times the AI assistant agent is right about its confidence ahead of time to handle some contingencies. Moreover, a	Objectives: 3, 4

[1] https://humansystems.arc.nasa.gov/groups/tlx/index.php
[2] See more recent works about design recommendations to create algorithms with a positive human-agent interaction and foster a pleasant user-experience: http://hdl.handle.net/1853/61232

	confusion matrix can be calculated to show: • True positive cases: AI assistant raises inaccuracy alarm indicating it has insufficient data to estimate the state of the grid and it actually doesn’t have the required data, • False positive cases: AI assistant raises inaccuracy alarm indicating it has insufficient data to estimate the state of the grid but it actually does have the required data (i.e. it should be confident but it isn't) • False negative cases: AI assistant doesn’t raise inaccuracy alarm, but is reality it can’t properly assess the situation (i.e. is falsely confident) *Note: This KPI is the equivalent of the “Assistant alert accuracy” KPI initially defined for UC1 “Power Grid Assistant”*	

1.6 Standardization opportunities and requirements

Classification Information
Relation to existing standards
ISO/IEC 23894:2023, Information technology — Artificial intelligence — Guidance on risk management. Operating the power grid is a high-stakes task, and therefore, risk management specifically related to AI is fundamental. This standard describes the principles applied to AI, risk management framework, and processes. It is intended to be used in connection (i.e., provides additional guidance for AI) with *ISO 31000:2018, Risk management – Guidelines*. *ISO/IEC 24029-2:2023, Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for using formal methods*. Artificial neural networks are generally a building block of AI assistants for power grid operation (see results from L2RPN competitions); thus, methodologies for using formal methods to assess the robustness properties of neural networks are important. This standard is focused on how to select, apply, and manage formal methods to prove robustness properties. The technical report ISO/IEC TR 24029-1:2021 complements this standard and presents an overview of different methods to assess the robustness of neural networks. *ISO/IEC 42001:2023, Information technology – Artificial intelligence – Management system*. This standard is the world’s first AI management system standard, providing valuable guidance for this rapidly changing field of technology. It addresses the unique challenges AI poses, such as ethical considerations, transparency, and continuous learning. For organizations, it sets out a structured way to manage risks and opportunities associated with AI, balancing innovation with governance. *IEEE 7000-2021, IEEE Standard Model Process for Addressing Ethical Concerns during System Design*. This standard defines a framework for organizations to embed ethical considerations in concept exploration and development. It promotes collaboration between key stakeholders and ensures ethical values are traceable throughout the design process, impacting the operational concept, value propositions, and risk management. It is applicable to all organizations, regardless of size or life cycle model.
Standardization requirements
Assessment of AI robustness should go beyond artificial neural networks (ISO/IEC 24029-2:2023) and consider other AI models, as well as the communication of this information to the end-user/decision-maker and the interaction between AI and the environment.

1.8 Societal concerns

Societal concerns
Description

Responsibility: Provide the capacity to evaluate the quality of the AI decisions and corresponding impacts in case of low-quality decisions and provide mitigation mechanisms to ensure security, integrity, validity, and accuracy of the AI assistant. **Explainability and transparency**: Disclose to stakeholders the evaluation methods used to assess robustness, explain AI failures (e.g., impact of input data contamination, communications failure), and allow them to submit test cases and adversarial examples. **Accountability:** Mitigate, detect, and correct erroneous or harmful AI decisions when the model is in operation. **Safety and security:** The AI system should perform consistently across different scenarios and considering the complexity of the environment in which the AI system will be used. The key question is to understand if technology is fit for its purpose and real-world operating conditions.
Sustainable Development Goals (SGD) to be achieved
SGD7. Affordable and clean energy / SGD13. Climate action

3 Step-by-step analysis of use case

3.1 Overview of scenarios

Scenario conditions					
No.	**Scenario name**	**Scenario description**	**Triggering event**	**Pre-condition**	**Post-condition**
1	Adaptation to real-world condition	The AI assistant adapts to partially missing and/or noisy data.	There is a chance that the system security is not ensured at the forecasted horizon if no action is performed. Thus, the AI assistant proposes actions to the operator.	The AI assistant is continuously checking that the transmission grid security is ensured at the appropriate horizons (e.g. from few hours ahead down to 30-min ahead) when considering list of contingencies defined in the operational policies of the TSO., with the addition that noisy and missing data can occur.	The recommendations from the AI assistant makes the human operator aware of the sensitivity to uncertainty of recommended actions.
2	Unfeasibility of action recommendation	The transmission grid state can't be estimated properly by the AI assistant, which can't propose any action recommendation.	Forecasting of transmission grid state can't be performed by the AI-assistant because quality of input data is too low and/or proportion missing data is too high. The AI assistant is raising an inaccuracy alarm.	Issues, inconsistencies are present in the data, and data are also missing.	Associated impacts are made visible to the human operator by the AI assistant.
3	Ineffectiveness of recommended action	The effect of actions recommended by the AI assistant is challenged by unexpected events or dynamics.	The AI-assistant has provided one or several action recommendations. The human operator has assessed that proposed actions are not feasible, or the actions didn't have the expected consequences on the transmission grid's state.	One or several prerequisites needed to perform an action recommended by the AI-assistant are missing or have changed. *or* Real-time behavior of the transmission grid is significantly different from simulation.	The AI-assistant proposes new alternative actions with the help of information provided by the human operator.
4	Changing grid	The transmission grid (or zone within a grid) is not the same as in model training phase.	See scenario 1 of UC.1 Power Grid.	See scenario 1 of UC.1 Power Grid.	See scenario 1 of UC.1 Power Grid.

Scenario conditions					
No.	Scenario name	Scenario description	Triggering event	Pre-condition	Post-condition
		The scenario can be decomposed in several levels of difficulty: • a grid element is added or removed on the zone • generation changes (e.g. increase of solar generation capacity) • the AI assistant is used on a different zone			

3 Requirements

Requirements		
Categories ID	**Category name for requirements**	**Category description**
Ro	Robustness	Encompasses both its technical robustness (ability of a system to maintain its level of performance under a variety of circumstances) as well as its robustness from a social perspective (ensuring that the AI system duly takes into account the context and environment in which the system operates). This is crucial to ensure that, even with good intentions, no unintentional harm can occur. *Source: EU-U.S. Terminology and Taxonomy for Artificial Intelligence. First Edition*
E	Efficiency	Ability of an AI system to achieve its goals or perform its tasks with optimal use of resources, including time, computational power, and data.
I	Interpretability	Make the behavior and predictions of AI systems understandable to humans, i.e., degree to which a human can understand the cause of a decision. *Source: Molnar, Christoph. Interpretable machine learning. Lulu. com, 2020.*
Requirement R-ID	**Requirement name**	**Requirement description**
Ro-1	Adaption to increased uncertainty	The AI-system should demonstrate the ability to sustain operational stability and decision performance in diverse and partially unpredictable scenarios, with increased forecasting errors.
Ro-2	Network change responsiveness	The AI system must be able to handle changes within the transmission grid infrastructure, such as the introduction of new grid elements and modifications in the grid topology as the Transmission grid evolves.
Ro-3	Cognitive load and stress	The AI system shall not increase the level of complexity of the situation, and associated level of stress for human operators (due to additional misinformation).
Ro-4	Reproducibility	All recommendations made by the AI system must be reproducible at a later point, given the same input. While the actions recommended by the system do not need to be identical in a strict mathematical sense - acknowledging the variability inherent in distributed computing environments - they should be closely aligned and functionally equivalent, ensuring reliable and predictable outcomes under similar conditions.
E-1	Computational efficiency	The AI system must be designed to ensure efficient training and inference capabilities on a range of computer hardware, from small-scale development setups with limited processing power to configurations involving multiple servers and multiple GPUs.
I-1	Level of interaction	Each operator has its own preferences (e.g. one operator can be more risk averse than others): ideally, the AI assistant interacting with a given operator could provide decision support that fits the preferences of this operator, but not necessary of another, especially given the type of situation that can require more attention. So the AI system shall be able to interact more or less with the human operator according to his/her preferences : • Fully manual • Get notified every time an overload is detected • Only get notified when AI assistant is not confident enough

4 Common Terms and Definitions

Common Terms and Definitions	
Term	**Definition**
TSO – Transmission System Operator	A natural or legal person responsible for operating, ensuring the maintenance of and, if necessary, developing the transmission

Common Terms and Definitions	
Term	**Definition**
	system in a given area and, where applicable, its interconnections with other systems, and for ensuring the long-term ability of the system to meet reasonable demands for the transmission of electricity. Source: Directive 2009/72/EC and ENTSOE glossary
EMS – Energy Management System	Optimal control center solution to enable secure, efficient, and optimized operation of the electric power system.
Contingency (“N-1” situation)	Electric system’s state after the loss of one grid element, and possibly several grid elements depending on the TSO’s policy
Load (or power) flow calculation	Calculations used to determine the voltage, current, and real and reactive power at various points in a power system under steady-state conditions.