

Disclaimer: This document serves as supplementary material to the deliverable report D1.1, titled “AI4REALNET Framework and Use Cases,” from the AI4REALNET project. It includes the responses to the ALTAI questionnaire for each use case within the project.

ALTAI questionnaire summary

Power Grid

REQUIREMENT #1 Human Oversight and Agency

Human Agency and Autonomy

Question	Decision	Consideration	Measure
<p>Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society?</p> <ul style="list-style-type: none"> • Could the AI system generate confusion for some or all end-users or subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision? • Are end-users or other subjects adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision? 	<p>Relevant (+)</p>	<p>The AI-assist is designed to provide recommendations to human operators in managing the power grid, which, in case of failure, might endanger the safety of property and people, lead to electricity outages and affect humans and economy.</p> <p>The AI can generate confusion via multiple actions and too many alarms when the AI is self-aware of the incapacity to solve the technical problem.</p> <p>The end-users are fully aware that the decision comes from an AI-based system.</p>	<p>This is covered by requirements Ro-2, Ro-5. However, note that the human operator is fully aware that is interacting with AI. In particular, it is not foreseen that AI systems have direct controls on grid elements:</p> <p>a) as input, they will receive information from environment and human operators,</p> <p>b) as output, they will provide information to human operator only. Human operator remains in charge for implementing actions.</p>

Question	Decision	Consideration	Measure
<p>Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system?</p> <ul style="list-style-type: none"> Are end-users or subjects informed that they are interacting with an AI system? 	Relevant (+)		<p>A requirement was added to the use case: Re-3 “Transparency to humans in terms of interaction with an AI system”.</p>
<p>Could the AI system affect human autonomy by generating over-reliance by end-users?</p> <ul style="list-style-type: none"> Did you put in place procedures to avoid that end-users over-rely on the AI system? 	Relevant (+)	<p>Humans remain in control, and AI acts as a recommendation system. However, with time, humans may start to trust more in AI, and there is the risk of over-reliance. Yet, the use case considers the need for alarms from the AI assistant as a functional requirement when it is not able to provide a recommendation that solves the technical problem, which contributes to mitigating over-reliance.</p>	<p>This is covered by requirements: Ro-2, Ro-4, I-1, I-3</p>
<p>Could the AI system affect human autonomy by interfering with the end-user’s decision-making process in any other unintended and undesirable way?</p> <ul style="list-style-type: none"> Did you put in place any procedure to avoid that the AI system inadvertently affects human autonomy? 	Relevant (-)	<p>AI acts as a recommender, therefore it will not impact human autonomy in a direct way.</p>	-
<p>Does the AI system simulate social interaction with or between end-users or subjects?</p>	Not Relevant (-)	No.	-

Question	Decision	Consideration	Measure
<p>Does the AI system risk creating human attachment, stimulating addictive behaviour, or manipulating user behaviour? Depending on which risks are possible or likely, please answer the questions below:</p> <ul style="list-style-type: none"> • Did you take measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System? • Did you take measures to minimise the risk of addiction? • Did you take measures to mitigate the risk of manipulation? 	<p>Relevant (+)</p>	<p>The assistant may create addictive behaviour in humans, but in normal conditions (i.e., without an adversarial attack to the output), it will not manipulate user behaviour.</p> <p><i>“Did you take measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System?”</i></p> <p>No, but this risk exists.</p> <p><i>“Did you take measures to minimise the risk of addiction?”</i></p> <p>No. The AI system is a recommender, and all the decisions are solely taken by the human operator. This somehow mitigates this risk.</p> <p><i>“Did you take measures to mitigate the risk of manipulation?”</i></p> <p>No. The output can be manipulated in case of a cyber-attack and the input as well. The input can also be affected by cyber-attacks or by gross errors/missing data (covered by use case 2).</p>	<p>Added as a new requirement HAO-1 “Mitigate addictive behaviour from humans”.</p>

Human Oversight

Question	Decision	Consideration	Measure
Please determine whether the AI system (choose as many as appropriate): <ul style="list-style-type: none"> • Is a self-learning or autonomous system. • Is overseen by a <i>Human-in-the-Loop</i>. • Is overseen by a <i>Human-on-the-Loop</i>. • Is overseen by a <i>Human-in-Command</i>. 	Relevant (+) Overseen by a Human-in-Command	The AI system provides recommendations which can be accepted or be adapted by the human at will.	-
Have the humans (human-in-the-loop, human-on-the-loop, human-in-command) been given specific training on how to exercise oversight?	Relevant (+)	Yes. The type of output is already known by the human (i.e., the same as traditional tools in power system control rooms), but humans should be trained to understand the rationale behind the AI system (e.g., understand how reinforcement learning works) and its limitations. Presently, a human operator does not have this training.	
Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?	Relevant (+)	Alarm is issued when the AI system is self-aware that it cannot produce a recommendation that solves the problem. Simulation (with a physically-based tool – power flow) is also possible to understand the impact of each recommendation in the system.	Requirement Ro-2 covers a confidence score which indicates potential undesirable recommendations
Did you ensure a 'stop button' or procedure to safely abort an operation when needed?	Not Relevant (-)	The AI system only works as a recommender. The human can override the AI system when necessary.	

Question	Decision	Consideration	Measure
Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?	Relevant (+)	An alarm is issued when the AI system is self-aware and cannot produce a recommendation that solves the problem. However, metrics that measure robustness offline and online are needed to monitor AI system performance.	Requirement Ro-2 and further metrics which will be developed

REQUIREMENT #2 Technical Robustness and Safety

Resilience to Attack and Security

Question	Decision	Consideration	Measure
Could the AI system have adversarial, critical or damaging effects (e.g., to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?	Relevant (+)	Yes, endanger the safety of property and people or outages in the electrical grid. This can occur due to different reasons: <ul style="list-style-type: none"> • Cyberattacks to input data, AI model output, and AI model • Noise and missing input data • High epistemic (model) uncertainty due to a lack of training data. 	Covered by requirements Ro-2 and Ro-8.
Is the AI system certified for cybersecurity (e.g., the certification scheme created by the Cybersecurity Act in Europe) or is it compliant with specific security standards?	Not Relevant (-)	No. However it is relevant, but outside the scope of AI4REALNET.	

Question	Decision	Consideration	Measure
<p>How exposed is the AI system to cyber-attacks?</p> <ul style="list-style-type: none"> • Did you assess potential forms of attacks to which the AI system could be vulnerable? • Did you consider different types of vulnerabilities and potential entry points for attacks such as: <ul style="list-style-type: none"> ○ Data poisoning ○ Model evasion ○ Model inversion 	Relevant (+)	<p>It can have cyberattacks to input data, model' output, and model. Moreover, cyber-attacks can occur during training and operation. This paper summarizes the potential attacks: Ilahi, I., Usama, M., Qadir, J., Janjua, M. U., Al-Fuqaha, A., Hoang, D. T., Niyato, D. (2021). Challenges and countermeasures for adversarial attacks on deep reinforcement learning. IEEE Transactions on Artificial Intelligence, 3(2), 90-109.</p> <p>These vulnerabilities can be studied in WP4 with a quantitative and qualitative analysis. Use case 2 is considering data "poisoning".</p>	Covered by requirements Ro-7 and Ro-8.
<p>Did you put measures in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?</p>	Relevant (+)	<p>No. However, this is important. Prior to these measures we should define first metrics (or KPIs) to monitor the robustness of the AI system during training and operation.</p>	Covered by requirements Ro-7 and Ro-8.
<p>Did you red team/pen test the system?</p>	Not Relevant (-)	<p>This is an important aspect to consider for a deliverable product, but not within the scope of this project. In particular it appears measures for pen testing will not benefit the project's development at an early stage.</p>	

Question	Decision	Consideration	Measure
<p>Did you inform end-users of the duration of security coverage and updates?</p> <ul style="list-style-type: none"> • What length is the expected timeframe within which you provide security updates for the AI system? 	<p>Relevant (+)</p>	<p>No. As mentioned above, this requires a continuous monitoring with adequate metrics of the AI performance.</p>	<p>Continuous monitoring of the performance metrics.</p>

General Safety

Question	Decision	Consideration	Measure
<p>Did you define risks, risk metrics and risk levels of the AI system in each specific use case?</p> <ul style="list-style-type: none"> • Did you put in place a process to continuously measure and assess risks? • Did you inform end-users and subjects of existing or potential risks? 	<p>Relevant (+)</p>	<p>No, but we should. However, we must distinguish two potential approaches:</p> <ul style="list-style-type: none"> • Risk-based approach where we quantify the probability and impact of a certain adversarial event. This may be challenging since we cannot make the analysis of the AI system separately from the physical environment. Somehow this resembles traditional reliability theory. • “Deterministic”-based approach where we use metrics that measure decrease in performance compared to a baseline scenario (or results from the training/learning stage). This is like quality control. <p>Informing end-users of potential risks can be made at the qualitative level, and also quantitative based on the training/learning stage.</p>	

Question	Decision	Consideration	Measure
<p>Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences?</p> <ul style="list-style-type: none"> • Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system? • Did you define safety criticality levels (e.g., related to human integrity) of the possible consequences of faults or misuse of the AI system? 	Relevant (+)	<p>One risk is the state vector (i.e., characterization of the operating context) that might have missing data, gross errors or even adversarial attacks, and this might lead to wrong decisions. This could be classified as an environmental threat, but another potential problem is that power grids are exposed to weather events. The risk of possible malicious attacks should be evaluated, but mainly at the state vector. Misuse of the AI system can lead to wrong decisions from the human operator.</p>	
<p>Did you assess the dependency of a critical AI system's decisions on its stable and reliable behaviour?</p> <ul style="list-style-type: none"> • Did you align the reliability/testing requirements to the appropriate levels of stability and reliability? 	Relevant (+)	Yes. Stability and reliability are essential in an AI assistant.	Requirements Ro-5 and Ro-7.
<p>Did you plan fault tolerance via, e.g., a duplicated system or another parallel system (AI-based or 'conventional')?</p>	Relevant (+)	No.	Requirement Ro-3.

Question	Decision	Consideration	Measure
Did you develop a mechanism to evaluate when the AI system has been changed to merit a new review of its technical robustness and safety?	Relevant (+)	No, but changes in the AI system should be auditable and controlled by humans. Nevertheless, several supervised and reinforcement learning algorithms have online learning and it might be difficult to make this evaluation or detection of change in the AI system. Thus, we should have automatic mechanisms to detect data shifts, but also AI model shifts.	Requirement Ro-9.

Accuracy

Question	Decision	Consideration	Measure
Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences?	Not Relevant (-)	In this concept of AI assistant, humans remain in control. The main consequence of low-level accuracy could be distrust from humans in AI and algorithmic aversion. If the AI actions were implemented automatically or with low human supervision, low accuracy could lead to critical consequences.	

Question	Decision	Consideration	Measure
Did you put in place measures to ensure that the data (including training data) used to develop the AI system is up to date, of high quality, complete and representative of the environment the system will be deployed in?	Relevant (+)	Yes. Note that the second use case for the power grids is about how to handle cases where the environment where the AI system was trained is different from the environment where it is running operationally. This means that transfer learning and adaptability of AI are important properties for deployment in real-world operating conditions.	Requirement E-4.

Question	Decision	Consideration	Measure
<p>Did you put in place a series of steps to monitor, and document the AI system's accuracy?</p>	<p>Relevant (+)</p>	<p>Continuous monitoring of the AI system is fundamental. It should be done at different levels:</p> <ul style="list-style-type: none"> • Measure performance continuously (online) with metrics such as reward score (objective function), human operator acceptance rate, alarms utility function, among others – KPIs defined in the use case. This performance can be quantified both during training and operational phases (e.g., identify changes in the environment compared to the training phase). • Conduct stress tests to assess robustness of the AI system, considering perturbation in the state vector (input data). These tests may also consider perturbations in the model (e.g., weights) and output. 	
<p>Did you consider whether the AI system's operation can invalidate the data or assumptions it was trained on, and how this might lead to adversarial effects?</p>	<p>Relevant (+)</p>	<p>Yes. This is the second use case.</p>	

Question	Decision	Consideration	Measure
Did you put processes in place to ensure that the level of accuracy of the AI system to be expected by end-users and/or subjects is properly communicated?	Relevant (+)	KPIs will be defined considering the two phases mentioned above.	Defining KPIs

Reliability, Fall-back plans and Reproducibility

Question	Decision	Consideration	Measure
<p>Could the AI system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility?</p> <ul style="list-style-type: none"> • Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals? • Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility? 	Relevant (+)	<p>Low reliability can recommend decisions that may not solve contingency problems and/or increase the risk of cascading effects. However, since it is an AI assistant, low reliability will mainly lead to low trust from human operators and not direct adversarial or damaging consequences. The AI system should be monitored permanently, but also its credibility and intimacy with the human operator. Reproducibility is important between the training and the operation environment, in particular the consistency of the decision quality. Use case 2 is about testing this reproducibility and also finding solutions to increase it.</p>	Requirement Ro-4.

Question	Decision	Consideration	Measure
<p>Did you put in place verification and validation methods and documentation (e.g., logging) to evaluate and ensure different aspects of the AI system's reliability and reproducibility?</p> <ul style="list-style-type: none"> Did you clearly document and operationalise processes for the testing and verification of the reliability and reproducibility of the AI system? 	Relevant (+)	Not yet, but this should be done in WP4, and it is an important requirement. However, we should define reliability and reproducibility since different interpretations are possible.	Development in WP4.
<p>Did you define tested failsafe fallback plans to address AI system errors of whatever origin and put governance procedures in place to trigger them?</p>	Relevant (+)	The human makes the final decision and can simulate the actions (i.e., run a power flow) provided by the AI assistant and check their quality. However, governance procedures should be defined to maintain the AI assistant in case of continuous poor performance.	Definition of governance procedures.
<p>Did you put in place a proper procedure for handling the cases where the AI system yields results with a low confidence score?</p>	Relevant (+)	Yes. The AI system generates alarms that inform the human operator about its confidence in solving technical problems considering past history and current context	Requirement Ro-2

Question	Decision	Consideration	Measure
<p>Is your AI system using (online) continual learning?</p> <ul style="list-style-type: none"> Did you consider potential negative consequences from the AI system learning novel or unusual methods to score well on its objective function? 	Relevant (+)	<p>Online learning is a possibility to deal with data shift. For instance, it is admissible when a new line or producer is added to the power grid, but offline training might be a better solution. Despite the time (available data) to learn new concepts, we do not see negative consequences from learning with new data.</p> <p>Online training can also be used for bidirectional learning (see first use case), where the AI assistant continuously learns from human actions and/or preferences.</p>	Requirements Ro-4 and Ro-5

REQUIREMENT #3 Privacy and Data Governance

Privacy

Question	Decision	Consideration	Measure
Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?	Not Relevant (-)	No. Privacy is not an issue.	
Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?	Not Relevant (-)	No. Privacy is not an issue.	

Data Governance

Question	Decision	Consideration	Measure
<p>Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?</p>	<p>Relevant (+)</p>	<p>No personal data is essential to be used during training. However, the AI system may leverage historical records of actions taken by human operators (i.e., imitation learning), but that can be fully anonymized since the operator's identification is not required. Yet, the action's timestamp is required, and when cross-referenced with a table of operator shifts, it may be possible to identify the operator and corresponding actions (and performance).</p>	<p>Requirement DG-1</p>

Question	Decision	Consideration	Measure
<p>Did you put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent?</p> <ul style="list-style-type: none"> • Data Protection Impact Assessment (DPIA); • Designate a Data Protection Officer (DPO) and include them at an early state in the development, procurement or use phase of the AI system; • Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications); • Measures to achieve privacy-by-design and default (e.g., encryption, pseudonymisation, aggregation, anonymization); • Data minimization, in particular personal data (including special categories of data); <p>Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system?</p> <p>Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?</p>	<p>Relevant (+)</p>	<p>This is covered by the Data Management Plan of the project. However, until now, no personal data is being collected.</p>	<p>Data management plan</p>

Question	Decision	Consideration	Measure
Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?	Relevant (+)	The AI system's input data (non-personal data) during training and operation deployment should be protected from the cyber-security viewpoint to avoid adversarial contamination or just gross/missing data from malfunctioning of the power grid metering system and SCADA.	Requirement Ro-6
Did you align the AI system with relevant standards (e.g., ISO, IEEE) or widely adopted protocols for (daily) data management and governance?	Not Relevant (-)	No.	

REQUIREMENT #4 Transparency

Traceability

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

<p>Did you put in place measures that address the traceability of the AI system during its entire lifecycle?</p> <ul style="list-style-type: none"> • Did you put in place measures to continuously assess the quality of the input data to the AI system? • Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)? • Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system? • Did you put in place measures to continuously assess the quality of the output(s) of the AI system? • Did you put adequate logging practices in place to record the decision(s) or recommendation(s) of the AI system? 	<p>Relevant (+)</p>	<p>Traceability is an important requirement for these use cases. In power grids, traceability is fundamental, and transmission system operators keep historical records of all main events. Thus, it is possible to replay scenarios where AI was used. Nevertheless, this means also storing the AI model (e.g., ANN weights, hyper-parameters) together with the input and output data.</p>	<p>Requirement Ro-4.</p>
---	---------------------	---	--------------------------

Explainability

Question	Decision	Consideration	Measure
<p>Did you explain the decision(s) of the AI system to the users?</p>	<p>Relevant (+)</p>	<p>Yes, explainability is an important target of the projects and the developments. However, in this stage most approaches rely on neural networks, where only feature importance (from sensitivity analysis or Shapley values) can be derived.</p>	<p>Requirement I-3</p>

<p>Do you continuously survey the users if they understand the decision(s) of the AI system?</p>	<p>Relevant (+)</p>	<p>Discussion: Explainability towards human operators is covered. But there are other stakeholders.</p> <p>We might go over the other stakeholders. à refine requirements for the other stakeholders</p> <p>Hypothesis: if we have an explanation for the human operator, then we can explain this to all other stakeholders (e.g., energy sector regulator).</p> <p>We may need to go beyond explainability (e.g., feature importance or feature-output modelling) with post-hoc methods and implement inherently interpretable models where the operator can understand how the models transform the input data into decisions. In this case, we may have a trade-off between accuracy and interpretability. An alternative is to apply knowledge distillation and use a second (and simpler model) to explain the decisions of the large</p>	
--	---------------------	--	--

Question	Decision	Consideration	Measure
		model.	

Communication

Question	Decision	Consideration	Measure
In cases of interactive AI systems (e.g., chatbots, robot-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?	Not Relevant (-)	In this case, the human operator is aware that an AI system is giving recommendations.	
<p>Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system?</p> <ul style="list-style-type: none"> • Did you communicate the benefits of the AI system to users? • Did you communicate the technical limitations and potential risks of the AI system to users, such as its level of accuracy and/ or error rates? • Did you provide appropriate training material and disclaimers to users on how to adequately use the AI system? 	Relevant (+)	<p>An alarm system (self-awareness) for the AI system is foreseen. The goal is to inform the user when the AI system may fail to solve the technical problem. This alarm can be generated with information about the operating context (using the input data/state vector as raw information) and the model uncertainty (epistemic uncertainty). Corresponding situations shall be evaluated in second use case (Sim2Real).</p> <p>The AI4REALNET digital environments can be used to train operators on how to use and interact with the AI system.</p>	Sim2Real, use case, Requirement Ro-2, AI4REALNET digital environments for training.

REQUIREMENT #5 Diversity, Non-discrimination and Fairness

Avoidance of Unfair Bias

Question	Decision	Consideration	Measure
<p>Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?</p>	<p>Relevant (+)</p>	<p>Proposed additional requirement:</p> <p>Category: Fa – Fairness – Ensure the recommendations and predictions of the AI system are in line with operating rules of TSOs, which in turn apply the principles of fairness, i.e., ensure fair distribution of strain/harm as well as fair access to the power grid (see applicable EU regulation covering TSOs’ operations).</p> <p>Requirement: Fa-1: The system must not unfairly favour specific producers of energy. A level playing field in the energy market as well as fair competition must be provisioned. Measures to ensure these fairness constraints are observed, must be put in place.</p> <p>In the discussion the two following issues regarding this proposition emerged:</p>	<p>Requirement FAIR-1.</p>

Question	Decision	Consideration	Measure
		<ul style="list-style-type: none"> • Occurring bias may very well originate from technical or physical limitations of electrical grid operations and hence may (in part or wholly) not avoidable. • Requiring the AI system to adhere to fairness standards that are not required from existing alternative techniques may put it at a disadvantage, especially if those originate from the source of the previous issue. 	

Question	Decision	Consideration	Measure
<p>Did you consider diversity and representativeness of end-users and/or subjects in the data?</p> <ul style="list-style-type: none"> • Did you test for specific target groups or problematic use cases? • Did you research and use publicly available technical tools, that are state-of- the-art, to improve your understanding of the data, model and performance? • Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system (e.g., biases due to possible limitations stemming from the composition of the used data sets (lack of diversity, non-representativeness)? • Where relevant, did you consider diversity and representativeness of end-users and or subjects in the data? 	<p>Relevant (+)</p>	<p>The AI assistant should be able to adapt to the operators’ individual features such as risk aversion, etc. to provide suitable recommendations.</p>	<p>Requirement I-4.</p>
<p>Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?</p>	<p>Not Relevant (-)</p>	<p>Not relevant for these use cases. Potentially addressed by using the ALTAI assessment.</p>	

Question	Decision	Consideration	Measure
<p>Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination, or poor performance of the AI system?</p> <ul style="list-style-type: none"> • Did you establish clear steps and ways of communicating on how and to whom such issues can be raised? • Did you identify the subjects that could potentially be (in)directly affected by the AI system, in addition to the (end-)users and/or subjects? 	<p>Relevant (+)</p>	<p>In these use cases, bias and discrimination can be directly towards certain grid users (generators, flexible loads) that are redispatched (or curtailed) more frequently than others. Using the physical equations of the power grid, it is possible to compare the decisions made by the AI system and the impact that other grid users would have in solving the technical problem. For instance, ex-post, it is possible to run an optimal power flow (OPF) with the redispatch costs and compare its solution with the AI system. Having a least-cost solution is the primary goal.</p>	<p>Verify operation at certain stages using OPF analysis.</p>

Question	Decision	Consideration	Measure
<p>Is your definition of fairness commonly used and implemented in any phase of the process of setting up the AI system?</p> <ul style="list-style-type: none"> • Did you consider other definitions of fairness before choosing this one? • Did you consult with the impacted communities about the correct definition of fairness, i.e., representatives of elderly persons or persons with disabilities? • Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness? • Did you establish mechanisms to ensure fairness in your AI system? 	<p>Relevant (+)</p>	<p>It is not used. No mechanisms exist to ensure fairness, but this should be included, although the definition is not trivial. In this domain, fairness is mainly used for the usage cost allocation of the transmission system among the network's users (generators and loads). However, metrics such as Jain's fairness index have been used to evaluate fairness in load shedding [1] and fairness in renewables' curtailment [2].</p> <p>[1] F. Moret and P. Pinson, "Energy Collectives: A Community and Fairness Based Approach to Future Electricity Markets," IEEE Trans. Power Syst., vol. 34, no. 5, pp. 3994–4004, Sep. 2019.</p> <p>[2] M. Z. Liu Liu, A. T. Procopiou, K. Petrou, L. F. Ochoa, T. Langstaff, J. Harding, and J. Theunissen, "On the Fairness of PV Curtailment Schemes in Residential Distribution Networks," IEEE Trans. Smart Grid, vol. 11, no. 5, pp. 4502–4512, 2020.</p>	<p>Requirement FAIR-2.</p>

Accessibility and Universal Design

Question	Decision	Consideration	Measure
Did you ensure that the AI system corresponds to the variety of preferences and abilities in society?	Not Relevant (-)	The power grid operation is concerned with providing electricity to its customers. This objective is not influenced by the variety of preferences and abilities in society.	
Did you assess whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion? <ul style="list-style-type: none"> • Did you ensure that information about, and the AI system's user interface of, the AI system is accessible and usable also to users of assistive technologies (such as screen readers)? • Did you involve or consult with end-users or subjects in need for assistive technology during the planning and development phase of the AI system? 	Not Relevant (-)	Not relevant for these use cases.	
Did you ensure that Universal Design principles are taken into account during every step of the planning and development process, if applicable?	Not Relevant (-)	Not clear which Universal Design principles we are talking.	

<p>Did you take the impact of the AI system on the potential end-users and/or subjects into account?</p> <ul style="list-style-type: none"> • Did you assess whether the team involved in building the AI system engaged with the possible target end-users and/or subjects? • Did you assess whether there could be groups who might be disproportionately affected by the outcomes of the AI system? • Did you assess the risk of the possible unfairness of the system onto the end-user's or subject's communities? 	<p>Relevant (+)</p>	<p>The AI will provide recommendations to human operators. It will not have a direct impact on target end-users and/or subjects. Most of this question was already answered before. Most notably the TSO Operators may be impacted by the AI system regarding their workload. The system shall support their work and not impede their performance, e.g., by increasing their workload.</p>	<p>KPI Workload</p>
--	---------------------	---	---------------------

Stakeholder Participation

Question	Decision	Consideration	Measure
<p>Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the AI system's design and development?</p>	<p>Relevant (+)</p>	<p>Stakeholders have been consulted during use case design and can be during the AI system design. Competitions with the digital environments will also help understand the AI system's benefits, limitations, and risks and extract lessons for further improvement.</p>	<p>Consultation of Stakeholders during the use case design. Conducting a public consultation.</p>

REQUIREMENT #6 Societal and Environmental Well-being

Environmental Well-being

Question	Decision	Consideration	Measure
Are there potential negative impacts of the AI system on the environment? <ul style="list-style-type: none"> Which potential impact(s) do you identify? 	Relevant (+)	The AI system will give priority to carbon-free actions, e.g., change network topology to avoid curtailment of renewable energy.	
Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)? <ul style="list-style-type: none"> Did you define measures to reduce the environmental impact of the AI system throughout its lifecycle? 	Relevant (+)	A KPI for carbon intensity is considered in the use case. The AI system will give priority to carbon-free actions. The AI system will also contribute to increasing resilience to extreme weather events and reduce the cost of blackouts.	KPI Carbon intensity

Impact on Work and Skills

Question	Decision	Consideration	Measure
Does the AI system impact human work and work arrangements?	Relevant (+)	The AI system will augment human operator analytics capabilities and decision-making tasks. It is not intended to replace the human.	HAO-1 and HAO-2
Did you pave the way for the introduction of the AI system in your organisation by informing and consulting with impacted workers and their representatives (trade unions, (European) work councils) in advance?	Not Relevant (-)	Human operators in control rooms already use supporting tools (mainly classical tools) to market and validate their decisions.	

Question	Decision	Consideration	Measure
Did you adopt measures to ensure that the impacts of the AI system on human work are well understood? <ul style="list-style-type: none"> Did you ensure that workers understand how the AI system operates, which capabilities it has and which it does not have? 	Relevant (+)	The proper use of data-driven tools requires training programs and risk assessment methodologies for humans and organizations.	HAO-2
Could the AI system create the risk of de-skilling of the workforce? <ul style="list-style-type: none"> Did you take measures to counteract de-skilling risks? 	Relevant (+)	Not really applicable at this stage: we must first characterize these risks, which will shall be done at a later stage in the project	HAO-2
Does the system promote or require new (digital) skills? <ul style="list-style-type: none"> Did you provide training opportunities and materials for re- and up-skilling? 	Not Relevant (-)	No need for new skills, but a higher knowledge about the fundamentals behind the AI system can help human operators to understand the decision support process.	

Impact on Society at large or Democracy

Question	Decision	Consideration	Measure
Could the AI system have a negative impact on society at large or democracy? <ul style="list-style-type: none"> Did you assess the societal impact of the AI system's use beyond the (end-)user and subject, such as potentially indirectly affected stakeholders or society at large? Did you take action to minimize potential societal harm of the AI system? Did you take measures that ensure that the AI 	Not Relevant (-)	Not relevant for these use cases.	

system does not negatively impact democracy?			
--	--	--	--

REQUIREMENT #7 Accountability

Auditability

Question	Decision	Consideration	Measure
Did you establish mechanisms that facilitate the AI system’s auditability (e.g., traceability of the development process, the sourcing of training data and the logging of the AI system’s processes, outcomes, positive and negative impact)?	Relevant (+)	Requirement: traceability of recommendation of the AI assistant down to the model. Saving the AI model (weights, hyperparameters, structure) is essential for auditability.	Requirement Ro-4.

<p>Did you ensure that the AI system can be audited by independent third parties?</p>	<p>Relevant (+)</p>	<p>Discussion:</p> <p>Audits are to be expected (though no formal process of assessment is available for software) → we need a req. for an audit.</p> <p>Audits have to be repeated (because data is used, so the development will be stochastic).</p> <p>If a customer has a complaint, the regulator will look at the case. We need to provide explanations (on a regular basis for each decision).</p> <p>AI Act will probably demand an audit (high-risk system). If the audit needs to be repeated often, is it necessary to automate the audit.</p>	<p>Requirement Acc-1.</p>
---	---------------------	---	---------------------------

Risk Management

Question	Decision	Consideration	Measure
<p>Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?</p> <ul style="list-style-type: none"> Does the involvement of these third parties go beyond the development phase? 	<p>Relevant (+)</p>	<p>During the development phase is unlikely. During the operational phase an audit might occur in case of outages or blackouts, or cyber-attacks to the input data.</p>	<p>Acc-1</p>

Question	Decision	Consideration	Measure
Did you organise risk training and, if so, does this also inform about the potential legal framework applicable to the AI system?	Relevant (+)	Risk and legal framework training will be important during design and operational phase.	
Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?	Not Relevant (-)	For these use cases it is not very relevant.	
Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)? <ul style="list-style-type: none"> • Does this process include identification and documentation of conflicts between the 6 aforementioned requirements or between different ethical principles and explanation of the 'trade-off' decisions made? • Did you provide appropriate training to those involved in such a process and does this also cover the legal framework applicable to the AI system? 	Relevant (+)	This is only relevant during the design and maintenance phases.	

Question	Decision	Consideration	Measure
<p>Did you establish a process for third parties (e.g., suppliers, end-users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?</p> <ul style="list-style-type: none"> Does this process foster revision of the risk management process? 	<p>Relevant (+)</p>	<p>This is fundamental, in particular, the creation of a database with vulnerabilities, risks and biases, similar to https://avidml.org/. However, the vulnerabilities and risks of other systems, e.g., SCADA, should be evaluated together due to interdependencies with the AI system (e.g., source of input data).</p>	<p>Requirement Acc-2.</p>
<p>For applications that can adversely affect individuals, have redress by design mechanisms been put in place?</p>	<p>Not Relevant (-)</p>	<p>For these use cases it is not very relevant.</p>	

Railway Network

REQUIREMENT #1 Human Agency and Oversight

Human Agency and Autonomy

Question	Decision	Consideration	Measure
<p>Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society?</p> <ul style="list-style-type: none"> - Could the AI system generate confusion for some or all end-users or subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision? - Are end-users or other subjects adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision? 	<p>Relevant (+)</p>	<p>The AI-assisted rescheduling is designed to provide support to human dispatchers. The resulting decisions affect the end-users (dispatchers) and society. Failure can result in increased workload for dispatchers and delays / cancellations for the public.</p>	<p>Addressed in I-2 (process transparency). Adequate training and education of end-users regarding the algorithmic nature of the system. This is out-of-scope for the proof-of-concept (POC) systems developed in this project.</p>
<p>Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system?</p> <ul style="list-style-type: none"> - Are end-users or subjects informed that they are interacting with an AI system? 	<p>Not Relevant (-)</p>	<p>The AI system is a tool, it should be designed as such and employees should receive appropriate training and information.</p>	

<p>Could the AI system affect human autonomy by generating over-reliance by end-users?</p> <ul style="list-style-type: none"> - Did you put in place procedures to avoid that end-users over-rely on the AI system? 	<p>Relevant (+)</p>	<p>(1) Automated AI-based rescheduling removes the human dispatcher from standard operations and thereby makes them reliant on the system.</p> <p>(2) AI-assisted rescheduling has the potential to generate over-reliance.</p>	<p>Addressed in I-2, I-3, I-4. Being able to adequately comprehend the systems choices will enable the dispatcher to avoid being overly reliant.</p> <p>(1)</p> <p>(2a) AI must expose its own biases and human operators must be aware of possible manipulations of their decision-making process and their own cognitive biases.</p> <p>(2b) Process which is closest to natural cognitive decision-making and is most likely to reduce over-reliance.</p> <p>(2c) Design human-machine interaction in such a way that the human perceives their role as a meaningful contribution to the joint decision-making process and can take responsibility for the decisions.</p>
--	---------------------	---	--

<p>Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way?</p> <ul style="list-style-type: none"> - Did you put in place any procedure to avoid that the AI system inadvertently affects human autonomy? 	<p>Relevant (+)</p>	<p>(1) No interference is possible given that the human is not actively making any decisions.</p> <p>(2) Yes, the recommendation system can influence human decision-making through effects such as the "anchor effect" or "confirmation bias".</p>	<p>(1) No measures required (2) No measures defined</p>
<p>Does the AI system simulate social interaction with or between end-users or subjects?</p>	<p>Not Relevant (-)</p>	<p>Neither system simulates social interaction, the system is designed to be used as a tool and does not emulate communication.</p>	

<p>Does the AI system risk creating human attachment, stimulating addictive behaviour, or manipulating user behaviour? Depending on which risks are possible or likely, please answer the questions below: I don't think human attachment could occur, given the AI system is planned as a tool. Manipulation of user behaviour might be possible?</p> <ul style="list-style-type: none"> - Did you take measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System? - Did you take measures to minimise the risk of addiction? - Did you take measures to mitigate the risk of manipulation? 	Not Relevant (-)	<p>Given that no social interaction is simulated, there is little to no risk of attachment to the system. Addiction and manipulation can be managed with adequate training and awareness.</p> <p>It is possible that the AI system manipulates human behaviour through the method of recommendation communication (i.e. confirmation bias and anchor effect).</p> <p>➔ These are discussed above</p>	
--	------------------	--	--

Human Oversight

Question	Decision	Consideration	Measure
<p>Please determine whether the AI system (choose as many as appropriate):</p> <ul style="list-style-type: none"> - Is a self-learning or autonomous system. - Is overseen by a <i>Human-in-the-Loop</i>. - Is overseen by a <i>Human-on-the-Loop</i>. - Is overseen by a <i>Human-in-Command</i>. 	(1) Self-Learning / Autonomous System, Human-on-the-Loop (2) Self-Learning / Autonomous System, Human-in-the-Loop, Human-in-Command		None required.
Have the humans (human-in-the-loop, human-on-the-loop, human-in-command) been given specific	Not Relevant (-)		Education and training are out of scope for this project.

training on how to exercise oversight?			
Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?	Not Relevant (-)		Such analyses and evaluation of the system is out of scope for this project, although preliminary suggestions and ideas can be given.
Did you ensure a 'stop button' or procedure to safely abort an operation when needed?	Relevant (+)	It is important for full control to transferrable to a human operator at all times. A transfer procedure must be developed.	(1) the human supervisor can take over control at any time. (2) the human agent is always in control and can ignore / shut down the AI assistant.
Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?	Relevant (+)	Online learning must be understood and interpretable to human agents.	Requirement I-4 added.

REQUIREMENT #2 Technical Robustness and Safety

Resilience to Attack and Security

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

<p>Could the AI system have adversarial, critical or damaging effects (e.g. to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?</p>	<p>Not Relevant (-)</p>		<p>Addressed by Ro-1, Ro-2, E-1.</p> <p>This project is limited to testing in simulation environments. Such considerations are therefore out-of-scope, albeit important for future work.</p> <p>Collision avoidance is managed by a separate system.</p>
<p>Is the AI system certified for cybersecurity (e.g. the certification scheme created by the Cybersecurity Act in Europe) or is it compliant with specific security standards?</p>	<p>Not Relevant (-)</p>	<p>It is relevant, but not within the general scope of AI4REALNET.</p>	<p>Re-1 considers compliance with regulations and standards</p>
<p>How exposed is the AI system to cyber-attacks?</p> <ul style="list-style-type: none"> - Did you assess potential forms of attacks to which the AI system could be vulnerable? - Did you consider different types of vulnerabilities and potential entry points for attacks such as: <ul style="list-style-type: none"> o Data poisoning o Model evasion o Model inversion 	<p>Not Relevant (-)</p>		<p>Cyber-attack risk assessment is out-of-scope, given that the developed systems are proof-of-concepts. Initial considerations of risk and potential weaknesses can be discussed.</p>
<p>Did you put measures in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?</p>	<p>Not Relevant (-)</p>		<p>Such measures are out-of-scope, given that the developed systems are proof-of-concepts.</p>

Did you red-team/pentest the system?	Not Relevant (-)	This is very important for a product to be delivered. For AI4REALNET the assessment of threats (as above) is planned.	Red-team and pentesting is out-of-scope, given that the developed systems are proof-of-concepts.
Did you inform end-users of the duration of security coverage and updates? - What length is the expected timeframe within which you provide security updates for the AI system?	Not Relevant (-)		Security coverage is out-of-scope, given that the developed systems are proof-of-concepts.

General Safety

Question	Decision	Consideration	Measure
Did you define risks, risk metrics and risk levels of the AI system in each specific use case? - Did you put in place a process to continuously measure and assess risks? - Did you inform end-users and subjects of existing or potential risks?	Not Relevant (-)	Risk assessment is important.	It is out-of-scope for this project. Such an analysis would need to occur upon implementation and be continuously monitored.

<p>Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences?</p> <ul style="list-style-type: none"> - Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system? - Did you define safety criticality levels (e.g. related to human integrity) of the possible consequences of faults or misuse of the AI system? 	<p>Not Relevant (-)</p>	<p>Initial analysis of potential risks can be done.</p>	<p>It is out-of-scope for a proof-of-concept system and would need to occur in detail upon implementation.</p>
<p>Did you assess the dependency of a critical AI system’s decisions on its stable and reliable behaviour?</p> <ul style="list-style-type: none"> - Did you align the reliability/testing requirements to the appropriate levels of stability and reliability? 	<p>?</p>		<p>Requirements Ro-1, Ro-2, E-1 and E-2.</p>
<p>Did you plan fault tolerance via, e.g. a duplicated system or another parallel system (AI-based or ‘conventional’)?</p>	<p>Relevant (+)</p>	<p>Given the criticality of the infrastructure managed by the system, fault tolerance is essential.</p>	<p>(1) the human operator can take over control upon system failure at any time.</p> <p>(2) The human operator is in control of the system and can turn off the AI assistant whenever necessary.</p>

<p>Did you develop a mechanism to evaluate when the AI system has been changed to merit a new review of its technical robustness and safety?</p>	<p>Relevant (+)</p>	<p>Performance monitoring is important, particularly when considering an online learning agent whose behaviour can change.</p>	<p>Processes consider post-hoc analysis of system performance (Step 8) and subsequent system adjustment by a human agent. Requirement Ro-3 ensures that quality can be evaluated post-hoc. Requirements I-1, I-2, I-3, and I-4 ensure that all processes (agent goals, option generation, decision-making and learning) are transparent and understandable.</p>
--	---------------------	--	---

Accuracy

Question	Decision	Consideration	Measure
<p>Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences?</p>	<p>Not Relevant (-)</p>	<p>For the recommender system the human will remain in full control, so no adverse consequences are foreseeable.</p> <p>For the automatic system low accuracy (that is not captures by human supervision) may lead to critical consequences such as FILL.</p>	<p>The system will be limited to operation in simulation environments, given that it is a proof-of-concept. This is therefore out-of-scope.</p>

Did you put in place measures to ensure that the data (including training data) used to develop the AI system is up-to-date, of high quality, complete and representative of the environment the system will be deployed in?	Not Relevant (-)		The developed systems are proof-of-concepts and trained in simulated environments. Preliminary ideas can be given, however this would need to be done before implementation.
Did you put in place a series of steps to monitor, and document the AI system's accuracy?	Relevant (+)	For development of the proof-of-concept systems, monitoring and documentation of system performance is essential.	The process considers monitoring and logging of system performance (Step 8), supported by requirements regarding transparency and interpretability (Ro-3, I-1, I-2, I-3, I-4)
Did you consider whether the AI system's operation can invalidate the data or assumptions it was trained on, and how this might lead to adversarial effects?	Relevant (+)	Online learning can lead to unforeseen changes in behaviour, necessitating explainability and interpretability requirements for the continuous learning process.	Requirement I-4 states that the online learning process is documented and interpretable for humans, allowing for continuous monitoring.
Did you put processes in place to ensure that the level of accuracy of the AI system to be expected by end-users and/or subjects is properly communicated?	Relevant (+)	System accuracy and uncertainty is a key element of interpretability and explainability, which is essential for human agents interacting with the AI system.	Requirement I-2 and I-3 cover process and decision explainability, which includes the communication of uncertainties.

Reliability, Fall-back plans and Reproducibility

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

<p>Could the AI system cause critical, adversarial, or damaging consequences (e.g. pertaining to human safety) in case of low reliability and/or reproducibility?</p> <ul style="list-style-type: none"> - Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals? - Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility? 	<p>Not Relevant (-)</p>		<p>Requirement Ro-3.</p> <p>The system is developed using simulations and is a proof-of-concept, therefore such considerations are out-of-scope.</p>
<p>Did you put in place verification and validation methods and documentation (e.g. logging) to evaluate and ensure different aspects of the AI system’s reliability and reproducibility?</p> <ul style="list-style-type: none"> - Did you clearly document and operationalise processes for the testing and verification of the reliability and reproducibility of the AI system? 	<p>Relevant (+)</p>	<p>Documentation and logging are essential for analysis of system performance.</p>	<p>Requirements I-4, Ro-3,</p> <p>Multiple existing requirements ensure that various system processes are logged and periodically reviewed by a human agent.</p>
<p>Did you define tested failsafe fallback plans to address AI system errors of whatever origin and put governance procedures in place to trigger them?</p>	<p>Relevant (+)</p>	<p>The criticality of the infrastructure managed by the AI systems requires such fallback plans / failsafes to be considered.</p>	<p>Both developed systems consider the human supervisor / agent to be the fallback plan and ensure that transition back to full human control is always possible.</p>

Did you put in place a proper procedure for handling the cases where the AI system yields results with a low confidence score?	Not Relevant (+)	Transparency and explainability inherently require that low certainty / accuracy is communicated and managed by the human agent. The exact procedure to follow is, however, out-of-scope and would need to be defined by the implementation partner.	Requirements I-2 and I-3 explicitly require that uncertainty / accuracy is communicated to the human agent.
Is your AI system using (online) continual learning? - Did you consider potential negative consequences from the AI system learning novel or unusual methods to score well on its objective function?	Relevant (+)	Online learning can lead to unforeseen changes in behaviour, necessitating explainability and interpretability requirements for the continuous learning process.	Requirement I-4 states that the online learning process is documented and interpretable for humans, allowing for continuous monitoring. Step 8 allows for human-triggered adjustments to the system.

REQUIREMENT #3 Privacy and Data Governance

Privacy

Question	Decision	Consideration	Measure
Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?	Relevant (+)	Important considerations, must be covered.	Re-1 ensures compliance with legal standards and regulation.
Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?	Not Relevant (-)	Such mechanisms are important.	They are out-of-scope for a proof-of-concept system.

Data Governance

Question	Decision	Consideration	Measure
Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?	Not Relevant (-)		Simulation environments are used for training.

<p>Did you put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent?</p> <ul style="list-style-type: none"> - Data Protection Impact Assessment (DPIA); - Designate a Data Protection Officer (DPO) and include them at an early state in the development, procurement or use phase of the AI system; - Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications); - Measures to achieve privacy-by-design and default (e.g. encryption, pseudonymisation, aggregation, anonymisation); - Data minimisation, in particular personal data (including special categories of data); - Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system? - Did you consider the privacy and data protection implications of data collected, 	<p>Not Relevant (-)</p>		<p>Such measures are out-of-scope for a proof-of-concept.</p>
---	-------------------------	--	---

generated or processed over the course of the AI system's life cycle?			
Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?	Not Relevant (-)		Data is generated in simulation environments, this is out-of-scope.
Did you align the AI system with relevant standards (e.g. ISO, IEEE) or widely adopted protocols for (daily) data management and governance?	Relevant (+)		Re-1 covers compliance with such standards.

REQUIREMENT #4 Transparency

Traceability

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

<p>Did you put in place measures that address the traceability of the AI system during its entire lifecycle?</p> <ul style="list-style-type: none"> - Did you put in place measures to continuously assess the quality of the input data to the AI system? - Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)? - Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system? - Did you put in place measures to continuously assess the quality of the output(s) of the AI system? - Did you put adequate logging practices in place to record the decision(s) or recommendation(s) of the AI system? 	Not Relevant (-)	Such measures are out-of-scope for a POC and therefore not considered.	I-2 considers the transparency of the decision-making process Ro-3 ensures retrospective quality control
---	------------------	--	---

Explainability

Question	Decision	Consideration	Measure
Did you explain the decision(s) of the AI system to the users?	Relevant (+)	Explainability and interpretability are essential for human operators and supervisors.	The 4 requirements I-1 to I-4 in the Interpretability category ensure that agent goals, option generation, decision-making and learning are transparent and understandable to the human agent.

Do you continuously survey the users if they understand the decision(s) of the AI system?	Not Relevant (-)		Such analyses are out-of-scope for a POC.
---	------------------	--	---

Communication

Question	Decision	Consideration	Measure
In cases of interactive AI systems (e.g., chatbots, robo-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?	Not Relevant (-)	The system is designed as a software tool, ensuring that it always clear to human agents that they are interacting with an artificial agent.	
Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system? <ul style="list-style-type: none"> - Did you communicate the benefits of the AI system to users? - Did you communicate the technical limitations and potential risks of the AI system to users, such as its level of accuracy and/ or error rates? - Did you provide appropriate training material and disclaimers to users on how to adequately use the AI system? 	Relevant (+)	Human operators and supervisors must understand the capabilities and limitations of the AI system to prevent misuse and foster trust.	Requirements I-1 to I-4 in the Interpretability category ensure that all aspects of the AI system (agent goals, option generation, decision-making, learning, capabilities, and limitation) are communicated to the human agent.

REQUIREMENT #5 Diversity, Non-discrimination and Fairness

Avoidance of Unfair Bias

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

<p>Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?</p>	<p>Relevant (+)</p>	<p>Bias avoidance must be considered during development and monitored after implementation.</p>	<p>Fairness requirements Fa-1 and Fa-2 introduced to ensure that the system fairly distributes unavoidable delays throughout the system and does not unfairly favour specific RUOMs.</p>
<p>Did you consider diversity and representativeness of end-users and/or subjects in the data?</p> <ul style="list-style-type: none"> - Did you test for specific target groups or problematic use cases? - Did you research and use publicly available technical tools, that are state-of- the-art, to improve your understanding of the data, model and performance? - Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system (e.g. biases due to possible limitations stemming from the composition of the used data sets (lack of diversity, non-representativeness)? - Where relevant, did you consider diversity and representativeness of end-users and or subjects in the data? 	<p>Not Relevant (-)</p>		<p>Analysis of end-user groups and diversity considerations cannot be done for a proof-of-concept.</p>
<p>Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?</p>	<p>?</p>	<p>Not relevant for these use cases. Potentially addressed by using the ALTAI assessment.</p>	

<p>Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination or poor performance of the AI system?</p> <ul style="list-style-type: none"> - Did you establish clear steps and ways of communicating on how and to whom such issues can be raised? - Did you identify the subjects that could potentially be (in)directly affected by the AI system, in addition to the (end-)users and/or subjects? 	<p>Not Relevant (-)</p>		<p>Such mechanism are out-of-scope for a POC.</p>
<p>Is your definition of fairness commonly used and implemented in any phase of the process of setting up the AI system?</p> <ul style="list-style-type: none"> - Did you consider other definitions of fairness before choosing this one? - Did you consult with the impacted communities about the correct definition of fairness, i.e. representatives of elderly persons or persons with disabilities? - Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness? - Did you establish mechanisms to ensure fairness in your AI system? 	<p>? Relevant (-)</p>	<p>An agreed upon definition of fairness is important.</p>	<p>Such a thorough analysis of fairness in the context of railway is out-of-scope for a POC. Develop a definition of fairness all partners agree upon.</p>

Accessibility and Universal Design

Question	Decision	Consideration	Measure
<p>Did you ensure that the AI system corresponds to</p>			

<p>the variety of preferences and abilities in society?</p>			
<p>Did you assess whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion?</p> <ul style="list-style-type: none"> - Did you ensure that information about, and the AI system's user interface of, the AI system is accessible and usable also to users of assistive technologies (such as screen readers)? - Did you involve or consult with end-users or subjects in need for assistive technology during the planning and development phase of the AI system? 			
<p>Did you ensure that Universal Design principles are taken into account during every step of the planning and development process, if applicable?</p>			
<p>Did you take the impact of the AI system on the potential end-users and/or subjects into account?</p> <ul style="list-style-type: none"> - Did you assess whether the team involved in building the AI system engaged with the possible target end-users and/or subjects? - Did you assess whether there could be groups who might be disproportionately affected by the outcomes of the AI system? - Did you assess the risk of the possible unfairness of the system onto the end-user's or subject's communities? 			

Stakeholder Participation

Question	Decision	Consideration	Measure
Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the AI system's design and development?	Relevant (+)	Incorporating stakeholders in the design process ensures that the developed systems fit the real-world requirements.	Stakeholder and public workshops conducted.

REQUIREMENT #6 Societal and Environmental Well-being

Environmental Well-being

Question	Decision	Consideration	Measure
Are there potential negative impacts of the AI system on the environment? - Which potential impact(s) do you identify?	Relevant (+)	We do not believe that the AI system will have any negative impacts on the environment but is more likely to have a positive impact through improved system efficiency.	
Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)? - Did you define measures to reduce the environmental impact of the AI system throughout its lifecycle?	Not Relevant (-)		Requirement O-2. Such mechanism are out-of-scope for a POC.

Impact on Work and Skills

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

Does the AI system impact human work and work arrangements?	Relevant (+)	Yes, the system does impact human work and work arrangements.	
Did you pave the way for the introduction of the AI system in your organisation by informing and consulting with impacted workers and their representatives (trade unions, (European) work councils) in advance?	Relevant (+)	Including the workers in the design process ensures that the developed systems are in line with their needs and expectations.	Workshops including workers.
Did you adopt measures to ensure that the impacts of the AI system on human work are well understood? <ul style="list-style-type: none"> - Did you ensure that workers understand how the AI system operates, which capabilities it has and which it does not have? 	Relevant (+)	It is important that the potential impacts of the developed systems are understood ahead of time, so that design considerations can be made during development.	Psychological considerations provided by the expertise of the FHNW group inform and guide the development process. Additionally, relevant KPIs are formulated.
Could the AI system create the risk of de-skilling of the workforce? <ul style="list-style-type: none"> - Did you take measures to counteract de-skilling risks? 	Relevant (+)	Yes, the AI system can cause de-skilling.	(1) no explicit measures mentioned. (2) this use-case lends itself to natural human decision-making and is designed to mitigate deskilling by actively involving the human decision-maker.
Does the system promote or require new (digital) skills? <ul style="list-style-type: none"> - Did you provide training opportunities and materials for re- and up-skilling? 	Relevant (+)	Yes, interaction with and supervision of the developed AI systems may require new skills (new type of decision-making, interaction with an artificial agent and a new system / interface introduced)	For both use-cases, design and realisation of training courses are out-of-scope for a POC, albeit essential for implementation.

Impact on Society at large or Democracy

Question	Decision	Consideration	Measure
<p>Could the AI system have a negative impact on society at large or democracy?</p> <ul style="list-style-type: none"> - Did you assess the societal impact of the AI system's use beyond the (end-)user and subject, such as potentially indirectly affected stakeholders or society at large? - Did you take action to minimize potential societal harm of the AI system? - Did you take measures that ensure that the AI system does not negatively impact democracy? 	Not Relevant (-)	Such considerations are out-of-scope for POC.	

REQUIREMENT #7 Accountability

Auditability

Question	Decision	Consideration	Measure
Did you establish mechanisms that facilitate the AI system's auditability (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact)?	Relevant (+)	Documentation and logging ensures auditability and is essential for post-hoc analyses and performance evaluations.	Requirement Ro-3 ensures retrospective quality control.
Did you ensure that the AI system can be audited by independent third parties?	Not Relevant (-)	Providing third parties with access to logged and documented data is out-of-scope for this project and would fall to the implementing entity.	

Risk Management

Question	Decision	Consideration	Measure
Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures? <ul style="list-style-type: none"> - Does the involvement of these third parties go beyond the development phase? 	Not Relevant (-)	Auditing is out-of-scope for a POC, but is made possible through documentation and logging.	
Did you organise risk training and, if so, does this also inform about the potential legal framework applicable to the AI system?	Not Relevant (-)	Risk analysis and training is out-of-scope for a POC.	
Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?	Not Relevant (-)	Establishing such ethics mechanisms is out-of-scope for a POC.	
Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)? <ul style="list-style-type: none"> - Does this process include identification and documentation of conflicts between the 6 aforementioned requirements or between different ethical principles and explanation of the 'trade-off' decisions made? - Did you provide appropriate training to those involved in such a process and does this also cover the legal framework applicable to the AI system? 	Not Relevant (-)	Monitoring processes are out-of-scope for a POC.	
Did you establish a process for third parties (e.g.	Not Relevant (-)	Reporting processes are out-of-	

<p>suppliers, end-users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?</p> <ul style="list-style-type: none"> - Does this process foster revision of the risk management process? 		<p>scope for a POC.</p>	
<p>For applications that can adversely affect individuals, have redress by design mechanisms been put in place?</p>	<p>Not Relevant (-)</p>	<p>A redress by design mechanism is out-of-scope for a POC.</p>	



Air Traffic Management

REQUIREMENT #1 Human Agency and Oversight

Human Agency and Autonomy

Question	Decision	Consideration	Measure
<p>Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society?</p> <ul style="list-style-type: none"> • Could the AI system generate confusion for some or all end-users or subjects on whether a decision, content, advice or outcome is the result of an algorithmic decision? • Are end-users or other subjects adequately made aware that a decision, content, advice or outcome is the result of an algorithmic decision? 	<p>Relevant (-)</p>	<ul style="list-style-type: none"> • Yes, if the proposed solution is not within the expected outcome but as the operator will have the final decision it has no operational impact. • Yes, unless the advisory is embedded on the ATM system and it is not clear its origin. Must be on a separate platform. <p>AF: AI system can generate confusion, if the generated prediction is not within the expected outcome. Still such situation cannot have an operational impact, as the operator will have the final decision. To assure, that the user understands the origin of the decision, it should be communicated on a separate platform.</p> <p>Is there a need for a requirement for a separate platform?</p>	<p>I-2</p>

Question	Decision	Consideration	Measure
<p>Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system?</p> <ul style="list-style-type: none"> Are end-users or subjects informed that they are interacting with an AI system? 	<p>Relevant (-)</p>	<p>Yes.</p> <p>AF: The end-users are aware, that they are interacting with an AI system, additionally the decision should be communicated through a separate platform for an additional visual reminder of the decision's origin (s. measure in the previous question).</p>	<p>I-2 (maybe extend the requirement to cover separate platform).</p>
<p>Could the AI system affect human autonomy by generating over-reliance by end-users?</p> <ul style="list-style-type: none"> Did you put in place procedures to avoid that end-users over-rely on the AI system? 	<p>Not relevant (?)</p>	<p>No. Not at this time.</p> <p>AF: Indirectly the use of assistant and the higher acceptance rate of decisions can with longer time affect the confidence and awareness of operators and reduce the time and effort they invest in checking the decisions generated by the assistant.</p>	<p>Partly I-2</p> <p>Can be indirectly assessed with KPIs: trust in AI solutions score, acceptance and agreement scores</p>

Question	Decision	Consideration	Measure
<p>Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way?</p> <ul style="list-style-type: none"> Did you put in place any procedure to avoid that the AI system inadvertently affects human autonomy? 	<p>Not relevant (?)</p>	<p>No. there is no interaction between operational system and AI system.</p> <p>AF: Low LoA (Level of Autonomy): the human autonomy is not affected. AI offers decisions for the human, which must be approved before being taken into action.</p> <p>High LoA: AI generated decisions are taken in action automatically; this affects human autonomy and demands stricter rules.</p>	<p>I-1, Ro-3</p> <p>Can be indirectly assessed with KPIs: trust in AI solutions score, acceptance and agreement scores</p>
<p>Does the AI system simulate social interaction with or between end-users or subjects?</p>	<p>Not Relevant</p>	<p>No.</p> <p>AF: Social interaction is not planned in the Use Cases.</p>	

Question	Decision	Consideration	Measure
<p>Does the AI system risk creating human attachment, stimulating addictive behaviour, or manipulating user behaviour? Depending on which risks are possible or likely, please answer the questions below:</p> <ul style="list-style-type: none"> • Did you take measures to deal with possible negative consequences for end-users or subjects in case they develop a disproportionate attachment to the AI System? • Did you take measures to minimise the risk of addiction? • Did you take measures to mitigate the risk of manipulation? 	<p>Not Relevant</p>	<ul style="list-style-type: none"> • No. Its not a chatbot type. • No. • No. <p>AF: No risk of addiction is expected according to the current description of use cases.</p>	

Human Oversight

Question	Decision	Consideration	Measure
<p>Please determine whether the AI system (choose as many as appropriate):</p> <ul style="list-style-type: none"> • Is a self-learning or autonomous system. • Is overseen by a <i>Human-in-the-Loop</i>. • Is overseen by a <i>Human-on-the-Loop</i>. • Is overseen by a <i>Human-in-Command</i>. 	Relevant (+)	<ul style="list-style-type: none"> • Overseen by Human-In-The-Loop, Human-in-Command <p>AF: The AI system provides recommendations which can be accepted or be adapted by the operator. At HIGH LOA with “management by exception” system becomes more autonomous.</p>	I-1
<p>Have the humans (human-in-the-loop, human-on-the-loop, human-in-command) been given specific training on how to exercise oversight?</p>	Not relevant (?)	<p>No.</p> <p>AF: Human operators should be trained for the new systems. There are currently no requirements for the development of studying materials and trainings.</p>	Partly O-1 (adapt to changing operational requirements)

Question	Decision	Consideration	Measure
<p>Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?</p>	<p>Not relevant (?)</p>	<p>No. The operator has the responsibility of assessing the AI outcome.</p> <p>AF: Alarm is issued when the AI system is self-aware that it cannot produce a recommendation that solves the problem.</p> <p>LOW LoA: operator checks each decision and only implements it if it is considered safe.</p> <p>Requirements can be derived from KPIs: system reliability, AI prediction robustness</p> <p>HIGH LoA: the automatically implemented decisions should be checked before they are taken into action to assure there are no undesirable effects.</p>	<p>Requirement Ro-3 demands reliable operation of the system.</p>

Question	Decision	Consideration	Measure
<p>Did you ensure a 'stop button' or procedure to safely abort an operation when needed?</p>	<p>Relevant (-)</p>	<p>No. To be assessed in the future.</p> <p>AF: LOW LoA - the AI system only works as a recommender. The human can correct or abort the decision when necessary.</p> <p>For HIGH LoA a stop button is required.</p>	
<p>Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?</p>	<p>Not relevant (?)</p>	<p>No.</p> <p>AF: If the environmental changes occur, that can affect the validity of generated decision, an alarm is issued for the operator to review the decision. Self-learning is not included directly in requirements, but they can be derived from KPIs: AI co-learning capability, AI prediction robustness.</p>	<p>Ro-3, E-1, I-1</p>

REQUIREMENT #2 Technical Robustness and Safety

Resilience to Attack and Security

Question	Decision	Consideration	Measure
Could the AI system have adversarial, critical or damaging effects (e.g., to human or societal safety) in case of risks or threats such as design or technical faults, defects, outages, attacks, misuse, inappropriate or malicious use?	Not relevant (?)	No. AF: HIGH LOA - if the decision of the system leads to adversarial danger and is implemented without the need of human confirmation, this can lead to damage.	
Is the AI system certified for cybersecurity (e.g., the certification scheme created by the Cybersecurity Act in Europe) or is it compliant with specific security standards?	Not relevant (?)	No. AF: Ro-2 demands that the system is protected against unauthorized access, cyber threats, and data breaches.	Ro-2
How exposed is the AI system to cyber-attacks? <ul style="list-style-type: none"> • Did you assess potential forms of attacks to which the AI system could be vulnerable? • Did you consider different types of vulnerabilities and potential entry points for attacks such as: <ul style="list-style-type: none"> ○ Data poisoning ○ Model evasion ○ Model inversion 	Relevant (+)	<ul style="list-style-type: none"> • No. • Automated systems as well as AI systems are prone to cyber intrusion and data corruption in a similar way. AF: These types of AI systems are prone to cyber intrusion and data corruption. Ro-2 demands that the system is protected against unauthorized access, cyber threats, and data breaches.	Ro-2
Did you put measures in place to ensure the integrity, robustness and overall security of the AI system against potential attacks over its lifecycle?	Not relevant (?)	No. AF: Not yet? the requirements for the security are set.	Ro-2

Question	Decision	Consideration	Measure
Did you red team/pen test the system?	Not relevant (?)	No.	
Did you inform end-users of the duration of security coverage and updates? <ul style="list-style-type: none"> What length is the expected timeframe within which you provide security updates for the AI system? 	Not relevant	NA. AF: The safety requirements will be applied by design due to the development according to the requirements.	

General Safety

Question	Decision	Consideration	Measure
Did you define risks, risk metrics and risk levels of the AI system in each specific use case? <ul style="list-style-type: none"> Did you put in place a process to continuously measure and assess risks? Did you inform end-users and subjects of existing or potential risks? 	Not relevant (?)	No. AF: Risk evaluation is essential for the design of safety properties of the systems. After evaluation the risks should be included into training materials.	Indirectly Ro-1 to identify unexpected situations.

Question	Decision	Consideration	Measure
<p>Did you identify the possible threats to the AI system (design faults, technical faults, environmental threats) and the possible consequences?</p> <ul style="list-style-type: none"> • Did you assess the risk of possible malicious use, misuse or inappropriate use of the AI system? • Did you define safety criticality levels (e.g., related to human integrity) of the possible consequences of faults or misuse of the AI system? 	Not relevant (?)	<p>No.</p> <p>s.above</p>	
<p>Did you assess the dependency of a critical AI system's decisions on its stable and reliable behaviour?</p> <ul style="list-style-type: none"> • Did you align the reliability/testing requirements to the appropriate levels of stability and reliability? 	Not relevant (?)	<p>No.</p> <p>AF: Stability and reliability are essential in an AI assistant. Two robustness requirements specify, that the system should work correctly under normal and unexpected circumstances.</p>	Ro-1, Ro-3
<p>Did you plan fault tolerance via, e.g., a duplicated system or another parallel system (AI-based or 'conventional')?</p>	Not relevant (?)	No.	Indirectly in Ro-3, but not specified

Question	Decision	Consideration	Measure
<p>Did you develop a mechanism to evaluate when the AI system has been changed to merit a new review of its technical robustness and safety?</p>	<p>Not relevant (?)</p>	<p>No.</p> <p>AF: There should be logging for the changes of algorithm, especially because algorithms are based on online reinforcement learning.</p> <p>Operator must be able to understand, if the version of the system changed while they were working on a specific decision, in case the new version proposes another decision, which might confuse operator.</p>	<p>O-2, I-2</p>

Accuracy

Question	Decision	Consideration	Measure
<p>Could a low level of accuracy of the AI system result in critical, adversarial or damaging consequences?</p>	<p>Relevant (+)</p>	<p>Yes, could be adversarial in the case of a deficient performance of the human-in-the-loop operator</p> <p>AF: The answer should be aligned with the first part of ALTAI Requirement 2. LOW LoA - no. There is still a risk of adversarial consequences in the case of a deficient performance of the human-in-the-loop operator. HIGH LoA - with low human supervision, low accuracy could lead to critical consequences.</p>	<p>Ro-1, Ro-3</p>
<p>Did you put in place measures to ensure that the data (including training data) used to develop the AI system is up to date, of high quality, complete and representative of the environment the system will be deployed in?</p>	<p>Relevant (-)</p>	<p>Yes. The data is provided is always the updated data (daily data).</p> <p>AF: E-2 requires scalability, but there is no requirement directly for the use of only up-to-date information. It is mentioned in the Use Case description, that the data is coming from many sources and the data infrastructure is demanding, maybe a validation procedure is needed to ensure the use of correct data base.</p>	

Question	Decision	Consideration	Measure
Did you put in place a series of steps to monitor, and document the AI system's accuracy?	Not relevant (-)	No. AF: According to the Use Case description all decisions and interactions will be logged. However, logging is not mentioned in the requirements, only in the use case description.	Logging functionality.
Did you consider whether the AI system's operation can invalidate the data or assumptions it was trained on, and how this might lead to adversarial effects?	Relevant (-)	Yes. See KPI design. AF: Which KPI? Is a KPI enough, if it's not in the requirement?	

Question	Decision	Consideration	Measure
<p>Did you put processes in place to ensure that the level of accuracy of the AI system to be expected by end-users and/or subjects is properly communicated?</p>	<p>Not relevant (?)</p>	<p>No.</p> <p>AF: The KPIs acceptance and agreement score are based on the comparison of AI-generated suggestions and decision accepted by the operator and are good metrics to evaluate this question.</p> <p>Are these KPIs communicated to the operator and who from the end-user should be aware of the level of accuracy, operators or supervisors?</p> <p>Additional possible KPIs to communicate the quality of the system's decisions:</p> <p>Decision Support satisfaction, Efficiency score, Trust in AI solutions score</p>	<p>I-2 – effective communication of information.</p>

Reliability, Fall-back plans and Reproducibility

Question	Decision	Consideration	Measure
<p>Could the AI system cause critical, adversarial, or damaging consequences (e.g., pertaining to human safety) in case of low reliability and/or reproducibility?</p> <ul style="list-style-type: none"> • Did you put in place a well-defined process to monitor if the AI system is meeting the intended goals? • Did you test whether specific contexts or conditions need to be taken into account to ensure reproducibility? 	<p>Not relevant (?)</p>	<p>No.</p> <p>At LOW LoA human operator will monitor the decisions and not apply any of them, that has potential danger.</p> <p>At HIGH LoA automatically implemented decisions can lead to adversarial consequences.</p> <p>The AI system should be monitored permanently, but also its credibility and intimacy with the human operator.</p> <p>KPIs: Trust in AI solutions score, Decision Support satisfaction, Efficiency score</p>	<p>Ro-1</p>
<p>Did you put in place verification and validation methods and documentation (e.g., logging) to evaluate and ensure different aspects of the AI system’s reliability and reproducibility?</p> <ul style="list-style-type: none"> • Did you clearly document and operationalise processes for the testing and verification of the reliability and reproducibility of the AI system? 	<p>Not relevant (?)</p>	<p>No.</p> <p>AF: KPIs: Significance of human revisions, System Reliability, AI prediction robustness</p>	

Question	Decision	Consideration	Measure
<p>Did you define tested failsafe fallback plans to address AI system errors of whatever origin and put governance procedures in place to trigger them?</p>	<p>Relevant (-)</p>	<p>Will do.</p> <p>AF: LOW LoA: The human makes the final decision and can simulate the provided by the AI assistant and check their quality.</p> <p>HIGH LoA: safety checks should be applied before the automated application of the decision.</p> <p>Governance procedures should be developed to specify the conditions for fallback.</p>	
<p>Did you put in place a proper procedure for handling the cases where the AI system yields results with a low confidence score?</p>	<p>Not relevant (?)</p>	<p>No.</p> <p>AF: On LOW LoA human operator will assess the decision and its accuracy score.</p> <p>At HIGH LoA an automated accuracy check can be implemented before the decision can be set to action.</p>	<p>Suggestion: automated security check.</p>

Question	Decision	Consideration	Measure
<p>Is your AI system using (online) continual learning?</p> <ul style="list-style-type: none"> Did you consider potential negative consequences from the AI system learning novel or unusual methods to score well on its objective function? 	<p>Not relevant (?)</p>	<p>No.</p> <p>AF: Continual learning is a part of reinforcement learning algorithms and ensures, that the system is adapting to data shifts, traffic load changes and new conditions. As long as the changes of algorithm go through the testing phase before the productive system state is updated and the user is aware, if the software state changed, continuous learning should not have negative impact.</p> <p>Additionally, the AI assistant continuously learns from human actions and/or preferences, which is assessed by KPI AI Co-learning Capability.</p>	<p>E-1, E-2, O-1</p>

REQUIREMENT #3 Privacy and Data Governance

Privacy

Question	Decision	Consideration	Measure
<p>Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?</p>	<p>Not Relevant</p>	<p>No.</p>	

Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?	Not Relevant	No.	
---	--------------	-----	--

Data Governance

Question	Decision	Consideration	Measure
Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?	Not Relevant	No.	

<p>Did you put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent?</p> <ul style="list-style-type: none"> • Data Protection Impact Assessment (DPIA); • Designate a Data Protection Officer (DPO) and include them at an early state in the development, procurement or use phase of the AI system; • Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications); • Measures to achieve privacy-by-design and default (e.g., encryption, pseudonymisation, aggregation, anonymization); • Data minimization, in particular personal data (including special categories of data); <p>Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system?</p> <p>Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?</p>	<p>Not Relevant</p>	<p>No. The system will not record/process personal data.</p>	
--	---------------------	--	--

Question	Decision	Consideration	Measure
Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?	Not Relevant	No.	Requirement Ro-6
Did you align the AI system with relevant standards (e.g., ISO, IEEE) or widely adopted protocols for (daily) data management and governance?	Not Relevant	NA	

REQUIREMENT #4 Transparency

Traceability

Question	Decision	Consideration	Measure
----------	----------	---------------	---------

<p>Did you put in place measures that address the traceability of the AI system during its entire lifecycle?</p> <ul style="list-style-type: none"> • Did you put in place measures to continuously assess the quality of the input data to the AI system? • Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)? • Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system? • Did you put in place measures to continuously assess the quality of the output(s) of the AI system? • Did you put adequate logging practices in place to record the decision(s) or recommendation(s) of the AI system? 	<p>Relevant (-)</p>	<ul style="list-style-type: none"> • Desirable. • Yes. (?) • Yes. (?) • Desirable. • Will have to be done. A must have. » New requirement <p>AF: Currently, information is scattered over various ATM systems, It makes the oversight of the input data and their quality assessment even more important for the accuracy of the decisions. It is planned to log all human interventions into AI decisions. The logging can be extended by the documentation of input data, that were used to generate the decision.</p>	
---	---------------------	---	--

Explainability

Question	Decision	Consideration	Measure
<p>Did you explain the decision(s) of the AI system to the users?</p>	<p>Relevant (-)</p>	<p>Will be done</p> <p>AF: The human can consult additional information and explanations underpinning the AI's decisions on demand, which is expected to foster trust in and acceptance of the AI system.</p> <p>KPI Trust in AI solutions score describes the operator's confidence in the AI-generated solution, with and without the need for additional explanations. Evaluating the difference between the KPI with and without explanation can show, if the explanations are helpful.</p> <p>KPI Prompt demand rate shows, how often the operator needs additional explanation and what kind of explanations are used</p>	<p>I-1</p> <p>KPI Trust in AI solutions score</p>

Question	Decision	Consideration	Measure
Do you continuously survey the users if they understand the decision(s) of the AI system?	Relevant (+)	<p>Will do</p> <p>AF: It can be indirectly measured by how often the operator requires additional explanations, which is already done to assess KPI Prompt demand rate.</p> <p>A routine for continuous survey can be implemented as a part of assessment of human-system interaction and the result can be logged together with the implemented decision.</p>	<p>I-1</p> <p>KPI Prompt demand rate</p>

Communication

Question	Decision	Consideration	Measure
In cases of interactive AI systems (e.g., chatbots, robot-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?	Not Relevant	<p>NA</p> <p>AF: The human operator is aware that an AI system is giving recommendations. This is also expected to be visually communicated using a separate platform.</p>	

<p>Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system?</p> <ul style="list-style-type: none"> • Did you communicate the benefits of the AI system to users? • Did you communicate the technical limitations and potential risks of the AI system to users, such as its level of accuracy and/ or error rates? • Did you provide appropriate training material and disclaimers to users on how to adequately use the AI system? 	<p>Relevant (-)</p>	<p>A must-do</p> <p>AF: This information should be included into the process of training the operators about how to use and interact with the AI system.</p>	<p>Training procedures, their contents and materials should be established.</p>
---	---------------------	---	---

REQUIREMENT #5 Diversity, Non-discrimination and Fairness

Avoidance of Unfair Bias

Question	Decision	Consideration	Measure
<p>Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?</p>	<p>Not Relevant</p>	<p>TUD to answer</p> <p>AF: It is worth assessing, if the data base can provide the fundament for any biases or if there are some groups that</p>	

Question	Decision	Consideration	Measure
<p>Did you consider diversity and representativeness of end-users and/or subjects in the data?</p> <ul style="list-style-type: none"> • Did you test for specific target groups or problematic use cases? • Did you research and use publicly available technical tools, that are state-of- the-art, to improve your understanding of the data, model and performance? • Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system (e.g., biases due to possible limitations stemming from the composition of the used data sets (lack of diversity, non-representativeness)? • Where relevant, did you consider diversity and representativeness of end-users and or subjects in the data? 	<p>Not Relevant (-)</p>	<p>NA</p> <p>AF: The AI assistant should be able to adapt to the operators' individual features such as risk aversion, etc. to provide suitable recommendations.</p> <p>It should be assessed, if the algorithm can learn personal biases of the operators during co-learning process. In this way the biases can be injected and affect the whole system.</p>	
<p>Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?</p>	<p>Not Relevant</p>	<p>NA</p>	

Question	Decision	Consideration	Measure
<p>Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination, or poor performance of the AI system?</p> <ul style="list-style-type: none"> • Did you establish clear steps and ways of communicating on how and to whom such issues can be raised? • Did you identify the subjects that could potentially be (in)directly affected by the AI system, in addition to the (end-)users and/or subjects? 	<p>Not Relevant</p>	<p>NA</p> <p>AF: In these use cases, bias and discrimination can be directed towards certain group of users – airlines, long/short-flights, military/civil flights. It is possible to include into the algorithms the estimation of effect the decision has on different groups of users.</p>	
<p>Is your definition of fairness commonly used and implemented in any phase of the process of setting up the AI system?</p> <ul style="list-style-type: none"> • Did you consider other definitions of fairness before choosing this one? • Did you consult with the impacted communities about the correct definition of fairness, i.e., representatives of elderly persons or persons with disabilities? • Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness? • Did you establish mechanisms to ensure fairness in your AI system? 	<p>Not Relevant</p>	<p>NA</p> <p>AF: It is not used or required, but it should be evaluated.</p>	

Question	Decision	Consideration	Measure
Did you ensure that the AI system corresponds to the variety of preferences and abilities in society?	Not Relevant	NA	
<p>Did you assess whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion?</p> <ul style="list-style-type: none"> • Did you ensure that information about, and the AI system's user interface of, the AI system is accessible and usable also to users of assistive technologies (such as screen readers)? • Did you involve or consult with end-users or subjects in need for assistive technology during the planning and development phase of the AI system? 	Not Relevant	<p>NA</p> <p>AF: Are there operators with disabilities, that can disturb their use of the system?</p>	
Did you ensure that Universal Design principles are taken into account during every step of the planning and development process, if applicable?	Not Relevant		

<p>Did you take the impact of the AI system on the potential end-users and/or subjects into account?</p> <ul style="list-style-type: none"> • Did you assess whether the team involved in building the AI system engaged with the possible target end-users and/or subjects? • Did you assess whether there could be groups who might be disproportionately affected by the outcomes of the AI system? • Did you assess the risk of the possible unfairness of the system onto the end-user's or subject's communities? 	<p>Relevant (-)</p>	<ul style="list-style-type: none"> • Yes. • NA • NA <p>A must have » New Requirement</p> <p>AF: The AI will provide recommendations to human operators. It will not have a direct impact on target end-users and/or subjects.</p> <p>The ATC staff may be impacted by the AI system regarding their workload. While AI can optimise operations, it also changes the nature of work, requiring a shift in skills for human operators who now need to oversee and interact with advanced AI systems.</p> <p>The introduction of the AI system might lead to concerns about job displacement and the need for reskilling of ATC staff.</p>	<p>Operators will have trainings to ensure they learn the new skills needed for the new type of work they will be performing with the AI system.</p>
--	---------------------	--	--

Stakeholder Participation

Question	Decision	Consideration	Measure
<p>Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the AI system's design and</p>	<p>Relevant (-)</p>	<p>We've been already putting in place a few consultation mechanisms/workshop (advisory board)</p>	<p>Consultation mechanisms/workshop (advisory board)</p>

development?		<p>AF: Stakeholders have been consulted during use case design and can be during the AI system design. Surveying of operators will also help understand the AI system's benefits, limitations, and risks and extract lessons for further improvement.</p>	
--------------	--	---	--

REQUIREMENT #6 Societal and Environmental Well-being

Environmental Well-being

Question	Decision	Consideration	Measure
<p>Are there potential negative impacts of the AI system on the environment?</p> <ul style="list-style-type: none"> Which potential impact(s) do you identify? 	Not relevant	<p>None at all</p> <p>The system aims to reduce the load on the air traffic system and reduce the environmental impact.</p>	
<p>Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?</p> <ul style="list-style-type: none"> Did you define measures to reduce the environmental impact of the AI system throughout its lifecycle? 	Not relevant (?)	<p>No</p> <p>AF: The optimization of air traffic management is expected to contribute to lower carbon emissions. The mechanisms can be determined to calculate the saved carbon emissions and the positive impact of the system.</p>	0-2

Impact on Work and Skills

Question	Decision	Consideration	Measure
Does the AI system impact human work and work arrangements?	Relevant (-)	<p>Yes. It is supposed.</p> <p>AF: The AI system will augment human operator analytics capabilities and decision-making tasks.</p>	
Did you pave the way for the introduction of the AI system in your organisation by informing and consulting with impacted workers and their representatives (trade unions, (European) work councils) in advance?	Not relevant (?)	<p>No</p> <p>AF: The trainings prior to implementation of the new AI system should help operator to overcome their doubts or fears concerning the change in their work methods.</p>	Operator trainings
<p>Did you adopt measures to ensure that the impacts of the AI system on human work are well understood?</p> <ul style="list-style-type: none"> Did you ensure that workers understand how the AI system operates, which capabilities it has and which it does not have? 	Relevant (-)	<p>Will do</p> <p>AF: The proper use of data-driven tools requires training programs and risk assessment methodologies for humans and organizations.</p>	Operator trainings
<p>Could the AI system create the risk of de-skilling of the workforce?</p> <ul style="list-style-type: none"> Did you take measures to counteract de-skilling risks? 	Relevant (-)	<p>Yes. It's a concern to be addressed.</p> <p>AF: The measures should be developed and conducted on a regular basis.</p>	Operator trainings

Question	Decision	Consideration	Measure
<p>Does the system promote or require new (digital) skills?</p> <ul style="list-style-type: none"> Did you provide training opportunities and materials for re- and up-skilling? 	Not relevant (-)	<p>Not directly for the end-user as it will be transparent</p> <p>AF: Extended knowledge about the fundamentals behind the AI system can help human operators to understand the decision support process. Trainings should be provided before the AI system is implemented.</p>	Operator trainings

Impact on Society at large or Democracy

Question	Decision	Consideration	Measure
<p>Could the AI system have a negative impact on society at large or democracy?</p> <ul style="list-style-type: none"> Did you assess the societal impact of the AI system's use beyond the (end-)user and subject, such as potentially indirectly affected stakeholders or society at large? Did you take action to minimize potential societal harm of the AI system? Did you take measures that ensure that the AI system does not negatively impact democracy? 	Not Relevant	NA	

REQUIREMENT #7 Accountability

Auditability

Question	Decision	Consideration	Measure
<p>Did you establish mechanisms that facilitate the AI system’s auditability (e.g., traceability of the development process, the sourcing of training data and the logging of the AI system’s processes, outcomes, positive and negative impact)?</p>	<p>Relevant (-)</p>	<p>(New) Req</p> <p>AF: Traceability of recommendation of the AI assistant down to the model. Saving the AI model (weights, hyperparameters, structure) and input data is essential for auditability.</p>	<p>Logging of each decision</p>
<p>Did you ensure that the AI system can be audited by independent third parties?</p>	<p>Relevant (-)</p>	<p>(New) Req</p> <p>AF: AI Act will probably demand an audit (high-risk system). If the audit needs to be repeated often, is it necessary to automate the audit.</p> <p>Additionally, due to the nature of RL algorithms they will change the state of the system and should be audited after each system update.</p>	<p>Define the system properties, that would be audited by the third parties and consider, if separate logging should be developed.</p>

Question	Decision	Consideration	Measure
<p>Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?</p> <ul style="list-style-type: none"> Does the involvement of these third parties go beyond the development phase? 	Not Relevant	Not now	
<p>Did you organise risk training and, if so, does this also inform about the potential legal framework applicable to the AI system?</p>	Not Relevant	<p>No.</p> <p>AF: Risk and legal framework training will be important during design and operational phase.</p>	
<p>Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?</p>	Relevant (-)	It's an idea to consider.	
<p>Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)?</p> <ul style="list-style-type: none"> Does this process include identification and documentation of conflicts between the 6 aforementioned requirements or between different ethical principles and explanation of the 'trade-off' decisions made? Did you provide appropriate training to those involved in such a process and does this also cover the legal framework applicable to the AI system? 	Not relevant (?)	<p>No.</p> <p>AF: Due to the nature of RL algorithms, they will be continuously updated. It would be advisable to reassess the system after updates to make sure the trustworthiness is not affected by the system changes.</p>	

Question	Decision	Consideration	Measure
Did you establish a process for third parties (e.g., suppliers, end-users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system? <ul style="list-style-type: none"> Does this process foster revision of the risk management process? 	Not Relevant	No.	
For applications that can adversely affect individuals, have redress by design mechanisms been put in place?	Not Relevant (-)	No. For these use cases it is not relevant.	