
# AI for real-world network operation

## WP6 – Project management and coordination

D6.4 – Ethics and Data Management Plan V2

AI4REALNET has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101119527, and from the Swiss State Secretariat for Education, Research and Innovation (SERI).

## DOCUMENT INFORMATION

| DOCUMENT | D6.4 – Ethics and Data Management Plan V2 |
|---|---|
| TYPE | Report |
| DISTRIBUTION LEVEL | PU - Public |
| DUE DELIVERY DATE | 30/09/2025 |
| DATE OF DELIVERY | 25/09/2025 |
| VERSION | V1.0 |
| 1DELIVERABLE RESPONSIBLE | INESC TEC |
| AUTHOR (S) | João Aguiar Castro, Ricardo Bessa |
| OFFICIAL REVIEWER/s | Manuel Schneider (FLATLAND), Eduardo Vilches (UKASSEL) |

## DOCUMENT HISTORY

| VERSION | AUTHORS | DATE | CONTENT AND CHANGES |
|---|---|---|---|
| **0.1** | Ricardo Bessa | 22/04/2025 | Template creation |
| **0.2** | Ricardo Bessa et al. | 30/06/2025 | Ethics assessment part |
| **0.3** | João Aguiar Castro | 15/07/2025 | Data management part |
| **0.4** | Ricardo Bessa | 16/07/2025 | Version to be shared with the Ethics and Data Protection Committee |
| **0.5** | Ricardo Bessa | 04/09/2025 | Changes following the EDPC meeting. Version ready for internal review |
| **1.0** | Ricardo Bessa | 25/09/2025 | Final version after internal review |

## ACKNOWLEDGEMENTS


| NAME | PARTNER |
|---|---|
| **Vasco Dias** | INESC TEC |
| **Duarte Dias** | INESC TEC |
| **Mohamed Hassouna** | Fraunhofer IEE |
| **Eduardo Vilches** | UKASSEL |
| **Herke van Hoof** | UvA |
| **Clark Borst** | TU DELFT |
| **Bruno Lemetayer** | RTE |
| **Marcello Restelli** | POLIMI |
| **Manuel Schneider** | Flatland Association |
| **Ricardo Chavarriaga** | ZHAW |
| **Anna Fedorova** | ZHAW |


## DISCLAIMER

*This project is funded by the European Union and SERI. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union and SERI. Neither the European Union nor the granting authority can be held responsible for them.*

# SUMMARY

Effective management of research data is crucial to ensuring that AI4REALNET datasets remain findable, accessible, interoperable, and reusable (FAIR) throughout the entire project lifecycle and well into the future. The initial sections of the Ethics and Data Management Plan set out the key practices and strategies required to ensure that AI4REALNET datasets adhere to the FAIR principles. This report provides guidance on recommended data management approaches and offers an overview of the datasets (DS) expected to be generated throughout the project. Detailed information on each dataset is presented across various sections of the document and will be progressively updated as the datasets evolve over their lifecycle. The list below presents a summary of the DS described in this document.

| DATASETS | TYPE OF DATA | PARTNERS INVOLVED |
|---|---|---|
| **DS. 1.1 - Feedback from participants in the stakeholders' workshops** | Text responses collected from workshop participants, containing professional feedback and anonymized personal data. | RTE, TenneT, SBB, DB, NAV, TU Delft, INESC TEC |
| **DS. 1.2 - Datasets for the use cases and digital environments** | Numerical data generated from synthetic simulations of grid, rail, and air operation environments. | All |
| **DS. 1.3 - Interactive IA with order-agnostic architectures** | Synthetic numerical data from AI-agent interactions. | All |
| **DS. 2.1 - Experimental and operational data collection for personalized biomarkers** | Psychophysiological and psychological measurements from human participants, including pseudonymized personal data collected via wearable sensors and questionnaires. | INESC TEC |
| **DS. 2.2 - Socio-technical analysis with dispatcher and traffic controller - Interviews** | Interview transcripts containing anonymized personal data from professionals. | LiU, TU Delft, NAV |
| **DS. 2.3 - Socio-technical analysis with dispatcher and traffic controller – Observation** | Observational records involving anonymized data from human participants. | LiU, TU Delft, NAV |
| **DS. 2.4 - Socio-technical analysis with dispatcher and traffic controller – Questionnaire** | Survey results including categorical and textual responses. | LiU, TU Delft, NAV |
| **DS. 2.5 - Knowledge-assisted AI performance** | Synthetic interaction logs from knowledge-based AI agents operating in simulated environments. | UvA |
| **DS. 2.6 - Hierarchical RL in relational domains** | Numerical and relational data generated from reinforcement learning agents in hierarchical domains. | POLIMI |
| **DS. 3.1 - State-action space for AI agent interaction with the digital environment** | Structured numerical data representing interactions between AI and digital environments. | All |
| **DS. 4.1 – Test datasets for evaluation of AI technical performance** | Synthetic and real-world datasets representing operational scenarios aligned with defined use cases. | All |

| DATASETS | TYPE OF DATA | PARTNERS INVOLVED |
|---|---|---|
| **DS. 4.2 - Test results for evaluation of AI technical performance** | Numerical data capturing adversarial test results and technical performance metrics of AI systems. | All |
| **DS. 4.3 - Data of network operator interactions and feedback produced during Human-AI evaluation scenarios** | Multimodal data including physiological signals, video, and survey responses from human-AI interaction experiments. | RTE, TenneT, SBB, DB, NAV, TU Delft, INESC TEC, LiU, TU Delft, FHNW |
| **DS. 4.4 - Test datasets for AI functional testing** | Synthetic numerical datasets designed to test AI functionalities. | All |
| **DS.4.5 - Feedback collected regarding the adoption of AI** | Anonymized survey responses reflecting participants´ perceptions of AI systems. | RTE, TenneT, SBB, DB, NAV, TU Delft, FHNW |
| **DS. 4.6- Human-in-the-loop Validation Data for AI-Supported Air Traffic Management** | Experimental data from human-in-the-loop trials in air traffic management. | TU Delft, ZHAW |
| **DS. 5.1 - Stakeholders map and contacts** | Restricted personal contact data compiled for stakeholder engagement. | INESC TEC |
| **DS. 5.2 - Dataset provided for AI competitions** | Numerical datasets generated for use in AI competitions, based on synthetic task environments. | All |

This document comprises a final part fully dedicated to the assessment and oversight of AI-related ethical implications, considering the application of AI in critical infrastructures, and ensures alignment with the EU AI Act, emphasizing the principles of trustworthy AI through proactive assessment, risk mitigation, and ethics-by-design integration. Rather than conducting formal legal or ethical compliance assessments, AI4REALNET follows the European Commission's trustworthy AI framework using the ALTAI (Assessment List for Trustworthy Artificial Intelligence) tool. This approach embeds ethical analysis into the project lifecycle, addressing dimensions such as transparency, human agency, accountability, and robustness. The ALTAI methodology was employed early in the design phase (WP1) to identify ethical risks across three application domains. Dedicated workshops and structured assessments allowed project stakeholders to link societal values to concrete requirements. In particular:

- Power Grid. Key ALTAI requirements, in particular accountability, transparency, robustness, and fairness, were strongly represented. AI supports operator decisions, retains human oversight, and prioritizes environmental goals.
- Railway network. Although proof-of-concept work is limited to simulations, ALTAI dimensions like human oversight and societal well-being were acknowledged. Broader ethical issues are deferred to later stages.
- Air traffic management (ATM). Transparency is vital. The AI assistant must ensure explainability and traceability to support human-AI collaboration in dynamic environments.

AI4REALNET introduced a risk management framework based on ISO 31000 and ISO/IEC 23894 standards. This framework divides AI-related risks into four dimensions: likelihood, magnitude,

vulnerability, and exposure. It was applied through stakeholder input across all domains. The framework informs continuous monitoring and adaptive improvement of requirements throughout the system development process.

A robust methodology, guided by ISO/IEC 24029-2, was proposed to evaluate system resilience against perturbations such as adversarial inputs, sensor failures, and environmental shifts. Adversarial datasets simulate attack scenarios, and custom perturbation models are created for each domain. This framework covered key evaluation metrics like a) robustness with performance under stress and resistance to perturbations, b) resilience with recovery capabilities after perturbations, and c) reliability with operational consistency and ability to handle out-of-distribution data.

Furthermore, the human-AI partnership is central to AI4REALNET. The project assesses how AI systems integrate into human decision-making through six key evaluation objectives (WP4, Task 4.3):

- Decision quality: interaction dynamics between human operators and AI recommendations.
- Trust and acceptability: operator trust in AI systems and explainability of decisions.
- User experience: cognitive alignment, stress levels, and satisfaction.
- Learning curves: co-evolution of human and AI performance.
- Task allocation: role balance in human-AI teaming.
- Long-term impacts: reflections on deskilling, bias, and social consequences.

The evaluation tools include psychophysiological metrics and subjective feedback to ensure comprehensive analysis.

Finally, the ethical dimensions identified in WP1 are being operationalized in algorithm design and human-AI collaboration models. These include a) develop explainable, robust AI via physics-informed models, safe reinforcement learning, and cognitive-aware human machine interfaces (WP2), and b) design co-learning architectures supporting human-in-the-loop, co-learning, and autonomous modes while preserving meaningful human control (WP3).

# TABLE OF CONTENTS

# LIST OF FIGURES



# LIST OF TABLES

# ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| AI | Artificial Intelligence |
| ALTAI | Assessment List for Trustworthy Artificial Intelligence |
| CC | Creative Commons |
| DMP | Data Management Plan |
| DPO | Data Protection Officer |
| DOI | Digital Object Identifier |
| ISO | International Organization for Standardization |
| EDPC | Ethics and Data Protection Committee |
| FAIR | Findable, Accessible, Interoperable, Reusable |
| GDPR | General Data Protection Regulation |
| ISO | International Organization for Standardization |
| VPN | Virtual Private Network |
| WP | Work Package |

**TABLE 1 – ABBREVIATIONS AND ACRONYMS**

# 1. INTRODUCTION

This second version of the Ethics and Data Management Plan (DMP) for AI4REALNET updates information about the datasets that will be collected or generated throughout the project and describes how the data will be managed in line with ethical, legal, and FAIR (Findable, Accessible, Interoperable, Reusable) principles.

An internal template was developed and shared with parents to collect dataset information (Annex 1). This information is updated as the project progresses.

The DMP is designed to be a living and working document, updated by the INESC TEC´s data steward, João Aguiar Castro, hereafter referred to as the DMP manager. The DMP manager will interact directly with the project coordinator (Ricardo Bessa) regarding the DMP updates. Moreover, the DMP manager is a point of contact for the project partners for issues related to the DMP and is responsible for providing documents to be filled out by partners to support said updates. Each dataset has a person responsible for its management and the preferred point of contact for updates to the datasets to be reported in the DMP throughout the project.

To raise awareness of the DMP benefits, the DMP manager was responsible for a presentation to introduce the project partners to the DMP components in the project kick-off meeting, 11 October 2023 (Aguiar Castro, 2023). On 1 March 2024, the first meeting of the Ethics and Data Protection Committee (EDPC) took place, the Data Protection Officer (DPO) and Ethics Advisors of each partner analyzed a draft version of the DMP and provided their recommendations for improvement in the document, and also some general suggestions to the project (Bessa, 2024). The meeting minutes can be found in Deliverable D6.2 (version V1 of the DMP). On 4 September 2025, a second meeting of the EDPC took place, and the feedback is presented in Section 6.6 of this document.

The DMP from version V1 has been expanded to include a dedicated section documenting the consortium's progress in assessing ethics compliance. This assessment aligns with both the ALTAI framework and the requirements of the AI Act. The new section clearly references the relevant project Work Packages (WPs) and deliverables that address robustness and risk analysis. These outputs reflect the active contribution of a multidisciplinary team, ensuring a comprehensive and responsible approach to ethical AI development.

This document is structured as follows:

- Data Summary (Section 2) presents a list of datasets expected to be produced in AI4REALNET, including their purpose and who is responsible for managing them.
- FAIR Data (Section 3) outlines practices for data documentation, sharing, citation, repository selection, and file naming.
- Data Security (Section 4) describes general data storage and backup practices that partners can apply locally.
- Legal Compliance (Section 5) presents general principles for handling data in accordance with the General Data Protection Regulation (GDPR) and describes how informed consent and anonymization are addressed.
- Ethics Compliance Assessment (Section 5) explains how the project follows an ethics-by-design approach through the ALTAI framework and aligns its activities with the values and requirements of the EU AI Act.

# 2. DATA SUMMARY

AI4REALNET will create datasets with a variety of typologies and requirements. This section will be updated continuously to further detail the datasets' information throughout their lifecycle based on the DMP information template.

This overview provides a general description of the types of datasets AI4RELANET will generate, their characteristics, and how they serve the project´s broader objectives.

To support the development, training, and validation of AI-based solutions across electricity, railway, and air traffic domains, AI4REALNET will produce a diverse collection of datasets. These include large-scale synthetic time series, AI agent interaction logs, operator psychophysiological signals, interviews, surveys, and human-in-the-loop experimental results. The datasets span both quantitative and qualitative formats, are sourced from simulated and real-world-inspired environments, and reflect realistic operating conditions, including edge cases and adversarial scenarios.

This diversity ensures that the AI framework under development is evaluated across a broad range of challenges: technical scalability, safety under uncertainty, human interaction, trustworthiness, and social acceptance. Data will be generated using digital environments (e.g., Grid2Op, Flatland, BlueSky), proprietary data collection tools, and experimental protocols, and will be used by various partners throughout the project lifecycle. Where applicable, ethical and legal considerations, including anonymization and data protection, will be addressed as part of the dataset's lifecycle.

The following tables provide an overview of the expected datasets by WP.

- Table 2 presents the Data Summary for *WP1 – Design of a holistic framework for AI in critical network infrastructures.*
- *WP2 - Fundamental AI building blocks* is covered in Table 3.
- The datasets associated with *WP3 - AI augmented human decision-making*, *WP4 - Validation and impact assessment* and *WP5 - Dissemination, communication, and exploitation of results*, are summarized in Table 4-Table 6 respectively.

Each dataset is described based on an assessment of its purpose, collection method, and contribution to the project´s objectives.

For completeness, a short description of the digital environments that will be used in the project, namely Grid2Op, Flatland, and BlueSky, is presented in Annex 2.

| Description of WP1 datasets |
| --- |
| **DS. 1.1 - Feedback from participants in the stakeholders' workshops.**<br>Workshops and consultations will be held with external stakeholders to collect feedback on the AI4REALNET use cases.<br><br>**Aim:** To validate the proposed AI framework with realistic and relevant use cases.<br>**How the data is collected:** Synthetic data will be generated by the Chronix2grid tool for electricity, tools for other domains shall be defined (generated by RTE for electricity, and collected by RTE as WP leader for the other domains).<br>**Type of data:** Text, photos, video, and sound recording.<br>**Software:** Online meeting recording tool and word processor software.<br>**Expected size and format:** GB-scale sound, text, and video.<br>**Conditions for sharing within the project:** The dataset can be shared without any constraint and at any time. |

| Description of WP1 datasets |
| --- |
| **Expected availability:** By the end of Task 1.2, it is not expected that this dataset will be made publicly available but only disclosed to the meeting´s participants. A summary and anonymized feedback will be included in Deliverable D1.1.<br>**Data Collection Period:** From 2024-24-01 to 2024-02-29 |
| **DS. 1.2 - Datasets for the use cases and digital environments**.<br>Each use case needs to be reflected with a corresponding dataset. Synthetic datasets are pre-build when installing a digital environment (Grid2op, Flatland, BlueSky), but additional synthetic data will be needed and generated to match the use cases and corresponding scenarios. This data will be used to evaluate how the developed solutions deal with real-world conditions (e.g., missing, wrong, or delayed measurements) and understand how the step from simulated environments to real-world environments impacts the performance of the solutions.<br><br>**Aim:** This dataset is related to the development of novel variants of supervised and reinforcement learning, supported by a set of functions and a human-machine interface for explainable, algorithmically transparent, and interpretable reinforcement learning.<br>**How the data is collected:** Synthetic data will be generated by the digital environments' functions: 1) ChroniX2grid tool in Grid2Op, which may use real-world operational data of the TENNET grid such as historical load and generation time series; 2) Flatland provides functions to generate synthetic railway networks and demand for trains with structural similarity to real networks, and samples from the German or from the Swiss railway network and schedules in anonymized form will be used to create problems that resemble parts of real railway networks; 3) BlueSky contains open data aircraft performance models, and an open global navigation database including 14480 airports, performance and operating procedure coefficients for 295 different aircraft types.<br>**Type of data:** Numerical<br>**Software:** Digital environments, Grid2op, Flatland, and BlueSky.<br>**Expected size and format:** GB scale CSV format<br>**Conditions for sharing within the project:** No specific conditions<br>**Expected availability:** The datasets will be part of the digital environments as pre-built data and/or embedded in the environment to generate additional synthetic data and scenarios for AI training and validation. Release in Deliverables D1.2-D1.4 (different software releases of the digital environment). |
| **DS. 1.3 - Interactive AI with order-agnostic architectures.**<br>This dataset contains the interaction of one or more AI agents with virtual environments such as Grid2Op or Flatland, including the states of the virtual environment, the actions selected by the AI agent, and any rewards obtained. One AI agent could make an arbitrary subset of decisions, after which the AI agent of interest is tasked with 'filling in' the remaining decisions.<br><br>**Note:** This data will also be generated in WPs 2-3 using the environments from WP1.<br><br>**Aim:** to provide a dataset for interactive (co-) learning. The dataset could be reused within and outside of the project to evaluate the capacity of learning algorithms to make good decisions under various imposed decisions by an interaction partner.<br>**How the data is collected:** Via API from the digital environment and an AI agent, and stored in a file or database.<br>**Type of data:** Numerical<br>**Software:** Log files will be compatible with freely available tools such as TensorBoard<br>**Expected size and format:** Logfile in, e.g., TensorBoard format, MB scale.<br>**Conditions for sharing within the project:** No specific conditions<br>**Expected availability:** With the publication of the relevant algorithms in WP2 deliverables, namely in the software releases of D3.2 and D3.3. |

**TABLE 2 – WP1 DESIGN OF A HOLISTIC FRAMEWORK FOR AI IN CRITICAL NETWORK INFRASTRUCTURES, DATA SUMMARY**

| Description of WP2 Datasets |
| --- |
| **DS. 2.1 - Experimental and operational data collection for operator personalized biomarkers.**<br>Make use of Psychophysiological data for human stress and cognition levels, implicit integration in digital environments, and AI agents to further make use of the data for enhanced interaction and empathy with the human operator. Experimental tests will be conducted using a validated experimental stress testing platform and a simple reaction time task procedure for cognitive performance analysis and personalization of algorithms. Surveys and biosensors will be used to collect data from human operators during this experimental protocol. After making use of this information for the personalization of the algorithms (based on Machine Learning techniques), the achieved models will be used in real-time to provide to the digital environment and the AI agents both stress and cognitive levels of the operator while it is performing tasks in the digital environment. This information will provide human context to the AI agents, complementing the operational context information that is already being provided.<br><br>**Ethical and legal issues:** Only anonymous and aggregated results may be disseminated or published in scientific publications (which may involve research teams from different institutions) with the consent of the participants (see Annex 4).<br><br>**Aim:** To understand how the user's psychophysiology changes can contribute to the digital environment's adaptation to the user during the various use cases. The dataset also aims to characterize each subject in the initial experimental test.<br>**How the data is collected:** Making use of proprietary wearable devices, paper questionnaires, and smartphones that store all the data locally and share it with the INESC TEC database server using a REST API. Data will be processed at INESC TEC servers, and output values (stress and cognitive levels) will be provided to project partners through the same REST API, using a secured protocol that complies with GDPR policies.<br>**Type of data:** Numerical (Discrete and continuous), categorical (nominal)<br>**Software:** Smartphone for real-time data acquisition (Android OS based application), Python and MATLAB for both data processing and model training (personalization), and SPSS for post-analysis of the data. Python is also used for model inference based on personalized models.<br>**Conditions for data sharing within the project:** Data collected for personalization purposes may be shared under an open-access policy, following full anonymization and data curation. Real-time outputs (e.g., stress and cognitive levels) will be made available to project partners via the secured REST API used for data transfer.<br>**Expected availability:** After experimental protocols and operational tests in specific use cases, the data needs to be curated and fully anonymized to be publicly available. A dataset might be prepared to be shared in the INESC TEC data repository by the end of WP2. |
| **DS. 2.2 - Socio-technical analysis with dispatcher and traffic controller – Interviews**.<br>This dataset is generated via qualitative data collection, such as interviews. This dataset aims to model the "work-as-done" process from which requirements for the technical and social system can be derived.<br><br>**Ethical or legal issues:** Personal or sensitive data will be handled based on informed consent. Only anonymized transcripts may be disseminated.<br><br>**Aim:** To model the "work-as-done" process from which requirements for the technical and social system can be derived.<br>**How the data is collected:** Interviews are conducted with dispatchers and traffic controllers.<br>**Type of data:** Audio record transcripts<br>**Software**: MAXQDA and audio recording<br>Expected size and format: mx22, mp3<br>**Conditions for sharing within the project:** Data can be shared within the project only after transcription and anonymization. Audio files will not be openly shared.<br>**Expected availability:** Only summarized and anonymized results will be included in Deliverables D2.3 and D4.3.<br>**Data Collection Period:** The interviews were held between 7 and 18 July 2025. |

| Description of WP2 Datasets |
| --- |
| **DS. 2.3 - Socio-technical analysis with dispatcher and traffic controller – Observation**<br>This dataset is generated via qualitative data collection, such as observation, with quantitative elements, such as questionnaires. This data set aims to model the "work-as-done" process from which requirements for the technical and social system can be derived.<br>**Ethical or legal issues:** As a non-intrusive method, observational data must still be handled ethically. Anonymization is mandatory for any sharing.<br>**Aim:** To collect information about the status quo concerning psychological criteria, such as motivation to identify technical and social system requirements.<br>**How the data is collected:** Direct observation during normal work activities, with structured note-taking.<br>**Type of data:** Text, description of observed subjects<br>**Software:** MAXQDA, MS Word<br>**Expected size and format:** mx22, docx<br>**Conditions for sharing within the project:** Data will only be shared after processing and anonymization.<br>**Expected availability:** Anonymized summaries will be shared only after anonymization processing.<br>**Data Collection Period:** The observations were held in June 2025. |
| **DS 2.4 - Socio-technical analysis with dispatcher and traffic controller – Questionnaire.**<br>This dataset is generated via quantitative data collection, such as questionnaires. This data set aims to collect information about the status quo concerning psychological criteria, such as motivation to identify technical and social system requirements.<br>**Ethical or legal issues:** Informed consent is required for data collection. Only aggregated and anonymized data will be disseminated.<br>**Aim:** To collect quantitative data regarding psychological criteria such as motivation to identify technical and social system requirements.<br>**How the data is collected:** Online questionnaire via the Tivian (Unipark) platform.<br>**Type of data:** text<br>**Software:** Tivian, SPSS, Excel<br>**Conditions for sharing within the project:** No specific conditions<br>**Expected availability:** Aggregated results will be included in Deliverables D2.3 and D4.3. External repository availability depends on reidentification risk analysis. |
| **DS. 2.5 - Knowledge-assisted AI performance.**<br>This dataset contains 1) the interaction of an AI agent with digital environments, including the states of the environment, the actions selected by the AI agent, and any rewards obtained, and 2) a description of any prior knowledge bases used by the AI agent.<br>**Aim:** to provide essential reinforcement learning tools that extend current capabilities in the project domains, which the other WP can use.<br>**How the data is collected: via** API from the digital environment and an AI agent and stored in a file or database.<br>**Type of data:** Numerical, relational<br>**Software:** Log files with freely available tools such as TensorBoard<br>**Expected size and format:** Logfile in, e.g., TensorBoard format, MB scale.<br>**Conditions for sharing within the project:** No specific conditions<br>**Expected availability:** With the publication of the relevant algorithms in WP2 deliverables, namely in the software releases of D3.2 and D3.3. |
| **DS. 2.6 - Hierarchical RL in relational domains.** |

| Description of WP2 Datasets |
| --- |
| This dataset contains the interaction of an AI agent with virtual environments such as Grid2Op or Flatland, including the states of the virtual environment, the actions selected by the AI agent, and any rewards obtained. In particular, the study of a hierarchical AI agent in a relational domain.<br><br>**Aim:** to provide basic RL tools that extend current capabilities in the project domains, which the other work packages can use.<br>**How the data is collected:** via API from the digital environment and an AI agent and stored in a file or database.<br>**Type of data:** Numerical, relational<br>**Software:** Log files will be compatible with freely available tools such as TensorBoard<br>**Expected size and format:** Logfile in, e.g., TensorBoard format, MB scale.<br>**Conditions for sharing within the project:** No specific conditions<br>**Expected availability:** With the publication of the relevant algorithms in WP2 deliverables, namely in the software releases of D3.2 and D3.3. |

**TABLE 3 – WP2 FUNDAMENTAL AI BUILDING BLOCKS, DATA SUMMARY**

| Description of WP3 datasets |
| --- |
| **DS. 3.1 - State-action space for AI agent interaction with the digital environment.**<br>This dataset is generated via interaction between an AI agent (e.g., a reinforcement learning agent) and the digital environment in a series of episodes. The digital environment generates the state data produced by considering synthetic or real data regarding network operating conditions, combined with a physical network (which can be based on a real network). The AI agent produces the action data and implements it in the digital environment. Moreover, a reward value is computed with a given reward function for each interaction. Datasets are generated using supervised ML from historical aviation traffic data and from interaction between an AI agent and the BlueSky simulation environment.<br><br>**Aim:** This dataset is related to the development of novel variants of supervised and reinforcement learning for large-scale complex networks, exploiting domain knowledge and hierarchical and distributed problem decomposition to increase scalability in realistic networks.<br>**How the data is collected:** via API from the digital environment and an AI agent and stored in a file or database.<br>**Type of data:** Numerical<br>**Software:** Grid2Op, Flatland, BlueSky. Log files will be compatible with freely available tools such as TensorBoard.<br>**Expected size and format:** MB scale. CSV or any file format, Logfile in, e.g., TensorBoard format, MB scale.<br>**Conditions for data sharing within the project:** No specific conditions<br>**Expected availability:** A curated set of interactions and results will be made available together with each WP3 deliverable, but earlier datasets can be published. |

**TABLE 4 – WP3 AI AUGMENTED HUMAN DECISION-MAKING, DATA SUMMARY**

| Description of WP4 datasets |
| --- |
| **DS. 4.1 - Test datasets for evaluation of AI technical performance.**<br>Datasets shall reflect realistic and representative operational scenarios, in accordance with the predefined use cases. Datasets shall also be scaled such that AI solutions' scalability can be evaluated.<br><br>**Aim:** To validate the proposed AI framework with realistic and relevant use cases.<br>**How the data is collected:** Synthetic data will be generated by the Chronix2grid tool for electricity, tools for other domains shall be defined (generated by RTE for electricity and collected by RTE as WP leader for the other domains).<br>**Type of data:** Numerical<br>**Software:** Digital environment.<br>**Expected size and format:** GB scale. CSV format. |

| Description of WP4 datasets |
| --- |
| **Conditions for sharing within the project:** The datasets can be shared without any constraint and at any time.<br>**Expected availability:** By the end of task 1.2, it is not expected that this dataset will be made publicly available but only disclosed to the meeting´s participants. Datasets shall be created at the beginning of WP4, and possibly updated during WP4 if needed.<br>**Data Collection Period:** The Dataset will be prepared from September 2025 to September 2026 (approximately). |
| **DS 4.2 - Test results from the evaluation of AI technical performance.**<br>This data will contain all results from tests carried out during Task 4.1, including the calculated KPIs.<br><br>**Aim:** To validate the proposed AI framework with realistic and relevant use cases.<br>**How the data is collected:** Via API or log data from the digital environment Grid2op for electricity – tools for other domains shall be defined, during and after test running, and stored in a file or database (generated by RTE for electricity and collected by RTE as WP leader for the other domains).<br><br>**Type of data:** Numerical<br><br>**Software:** Any numerical data processor software.<br><br>**Expected size and format**: GB-scale CSV format<br>**Conditions for sharing within the project**: The dataset can be shared without any constraint and at any time.<br>**Expected availability:** During task 4.1, to be published as part of the work report from the project.<br>**Data Collection Period:** From 1st October to 31st March for the 1st validation campaign; From 1st April 2026 to 31st March for the 2nd validation campaign. |
| **DS 4.3 - Data on network operator interactions and feedback produced during Human-AI evaluation scenarios.**<br>Experimental protocols will be conducted to evaluate balance between AI and human in the selected use case. This means that biosensors will be used to collect psychophysiological measures that provide information on human operator internal processes, user experience and acceptability along several dimensions (e.g., attention, mental workload, stress). These data will be collected from operational staff and experts from TAP, SBB, DB, TenneT and RTE involved in AI4REALNET project.<br><br>**Ethical or legal issues:** Shall be investigated since data is generated following the intervention of humans or influence in a physical environment, in any case the dataset must be anonymized.<br><br>**Aim:** To validate the proposed AI framework with realistic and relevant use cases.<br>**How is collected:** Surveys, live recording (generated by RTE since coming from its employees and collected by RTE as WP leader for the other domains).<br>**Type of data:** Text, psychophysiological measures, video (e.g., eye position tracking).<br>**Software**: Biosensors (for psychophysiological measures), cameras, and any word processor software.<br>**Expected size and format:** GB scale text and video.<br>**Conditions for sharing within the project:** If anonymized, the dataset can be shared without any constraint and at any time.<br>**Expected availability:** During task 4.3, to be published as part of the work report from the project (anonymized).<br>**Data Collection Period:** Data will be collected during the 2nd validation campaign from 1st April 2026 to 31st March 2 |
| **DS. 4.4 – Test datasets for AI functional testing.**<br>Functional testing will be conducted to validate AI (i.e., pre-trained AI agent) safety functionalities, considering the data-driven nature of AI and the fact that security and data breaches of AI systems can progressively influence future decision quality due to continuous training and updating of the underlying AI system. The goal of these adversarial datasets, generated by either a heuristic method with domain knowledge or a |

| Description of WP4 datasets |
| --- |
| reinforcement learning based algorithm, is to evaluate how the developed solutions deal with real-world conditions, such as missing, wrong, or delayed measurements, data distribution shift, etc., to understand how the step from simulated environments to real-world environments impacts the performance of the solutions.<br><br>**Aim:** To validate the proposed AI framework in various use cases and realistic digital environments for the operation of critical infrastructures, particularly considering the robustness of data-driven approaches in real operating conditions.<br>**How the data is collected:** Via API from the digital environment and AI agent, and stored in a file or database.<br>**Type of data:** Numerical<br>**Software:** The digital environments of the project (Grid2Op, Flatland, Bluesky)<br>**Expected size and format**: MB scale. CSV or npy file format.<br>**Conditions for sharing within the project: T**he dataset can be shared without any constraint and at any time, and the data can be generated with access to the Python code that can be integrated with each digital environment.<br>**Expected availability:** The software code for the functional testing and generation of adversarial datasets on the top of the data generation engines available in each environment will also be released. This should happen between the months of M30 and M42 in the project GitHub. |
| **DS. 4.5 - Feedback collected regarding the adoption of AI.**<br>Interviews will be conducted with industry experts involved in the AI4REALNET project to collect reactions (e.g., qualms, worries, and positive aspects) towards adopting AI under employment law and workers' proper perspective.<br><br>**Ethical or legal issues:** Shall be investigated since the data is generated following intervention of humans or influence in a physical environment, in any case this dataset must be anonymized.<br><br>**Aim:** To increase social, academic (AI community) and business awareness to AI potential.<br>**How the data is collected:** Surveys (generated by RTE since coming from its employees and collected by RTE as WP leader for the other domains).<br>**Type of data:** Text, Audio record and transcripts<br>**Software:** Any word processor software.<br>**Expected size and format:** MB scale, text format.<br>**Conditions for sharing within the project:** The dataset can be shared without any constraint and at any time.<br>**Expected availability:** During task 4.4, to be published as part of the report from the project (anonymised). One time survey.<br>**Data collection period:** Data will be collected during the 2nd validation campaign from 1st April 2026 to 31st March 2027. |
| **DS. 4.6 - Human-in-the-loop Validation Data for AI-Supported Air Traffic Management.**<br>This dataset results from human-in-the-loop evaluations conducted with air traffic controllers (ATCo) and staff managers, as part of the validation process for AI-support air traffic management solutions. It includes qualitative data collected through assessment questionnaires covering workload, trust, acceptance, and situational awareness. In addition, it comprises modified sector plans and routing structures created by ATCos while interacting with AI-generated solutions and the operational environment. On the quantitative side, the dataset includes traffic flow performance metrics, such as routing efficiency and safety indicators (e.g., minimum separation distances and convergence). It also contains the predicted ATCo taskload per sector, derived from interactions during the validation exercises.<br><br>**Ethical and legal issues:** To be assessed, as the dataset involves human participants (air traffic controllers and managers) and includes subjective input and operational data. Anonymization and appropriate safeguards must be ensured in compliance with GDPR and ethical standards. |

| Description of WP4 datasets |
|---|
| **Aim:** To validate the proposed AI framework using realistic and operationally relevant use cases. The focus includes validating multi-objective performance as well as the explainability, transparency, and interpretability of reinforcement learning (RL) methods developed.<br>**How the data is collected:** Data will be collected through a **human-in-the-loop experiment**, involving active participation of air traffic controllers and planners.<br>**Type of data:** numerical and text (questionnaires)<br>**Software:** BlueSky and Sector X<br>**Expected size and format:** MB scale binary format. CSV.<br>**Conditions for sharing within the project:** Data can be shared when anonymized<br>**Expected availability:** It will depend on approval from ATCo/planner consent.<br>**Data collection period:** WP4 timeline. Trials are expected to start in Q2-2026. |

**TABLE 5 – WP4 VALIDATION AND IMPACT ASSESSMENT, DATA SUMMARY**

| WP5 Datasets | Aim for Data Collection |
|---|---|
| **DS. 5.1 - Stakeholders' map and contacts.**<br>Contact list of project stakeholders in different domains (AI, energy, mobility, etc.) and their feedback collected during meetings, workshops, and events. This includes the External Experts Advisory Board.<br><br>**Aim:** to develop a strategic ecosystem value map, including the protocol for identifying and analyzing the stakeholders. Organization of ecosystem engagement events to share knowledge and to foster synergies between the different initiatives.<br>**How the data is collected:** Survey and questionnaires, newsletter subscription, and feedback during online meetings (video).<br>**Type of data:** E-mail, contact, text (name, company), video.<br>**Software:** LimeSurvey, webpage, MS Teams.<br>**Expected size and format:** MB scale, .xlsx format, .mp4 format.<br>**Conditions for sharing within the project**: No specific conditions. **Expected availability:** It is to be used only by the WP5 leader (INESC TEC). | |
| **DS 5.2 - Dataset provided for AI competitions.**<br>The dataset will include synthetic data, which are network models and time series to be used by competition participants.<br>An example of an AI challenge run by RTE in 2023 for the power grid is the l2rpn_idf_2023 dataset, which is a 12GB CSV file created on the IEEE 118 grid. It includes time series with loads, productions, and next-timestep forecasts for 500 years at 5-minute resolution (network, generator, demand, wind, and solar signals).<br><br>**Aim:** To increase social, academic (AI community), and business awareness of AI potential.<br>**How the data is collected: D**ata will be either taken from a past AI challenge (e.g. l2rpn_idf_2023 dataset) or created from Real-world operational data of the RTE's transmission grid (generated by RTE and TenneT).<br>**Type of data:** Numerical<br>**Software: a** digital environment where the solutions are evaluated (Grid2op for electricity).<br>**Expected size and format:** GB scale. CSV format.<br>**Conditions for sharing within the project:** The dataset can be shared without any constraints and at any time.<br>**Expected availability:** During task 5.2, to be published for the competition.<br>**Data Collection Period:** The dataset will be generated before the start of the competition, so between September 2025 and June 2026 (approximately). | |

**TABLE 6 – WP5 DISSEMINATION, COMMUNICATION, AND EXPLOITATION OF RESULTS, DATA SUMMARY**

# 3. FAIR DATA

This section describes the guiding principles for promoting the FAIRness of AI4REALNET data by describing high-level practices to be adopted during the project. While AI4REALNET embraces an approach that is *as open as possible,* data availability must be assessed on a case-by-case basis, considering the requirements that the distinct types of data may have.

AI4REALNET is committed to the following underlying principles:

- The data collected and generated during the project will be made available to project partners and made openly available concerning possible embargo periods through data repositories.
- Persistent identifiers are assigned to the data deposited in repositories and are included in the metadata.
- Metadata should be made accessible even when data is not publicly available.
- Metadata is based on standard vocabularies whenever possible, but customizable metadata may also be required.

## 3.1 DATA AVAILABILITY

AI4REALNET manages the availability of datasets according to ethical, legal, and technical constraints, as well as the intended role of each dataset within the project. Datasets vary in terms of their sensitivity, relevance for external dissemination, and degree of maturity. As such, three main categories are considered:

- **Openly available data**: datasets (or their corresponding tools/codebases) that are already available or scheduled for public release with a clear timeline.
- **Pending public availability**: datasets that are not yet available but are expected to be released after deliverable completion, curation, and anonymization.
- **Internal access only**: datasets intended exclusively for internal use within the consortium.

### 3.1.1 OPENLY AVAILABLE DATA

Open datasets contribute to transparency, reproducibility, and the reuse of project outputs. Table 7 will be updated as additional datasets are released in open format over the course of the project.

| Open Available Dataset |
| --- |
| **Structured Power Grid Simulation Dataset for Machine Learning: Failure and Survival Events in Grid2Op´s L2RPN WCCI 2022 Environment**<br><br>https://doi.org/10.5281/zenodo.13948340<br><br>**Note:** This dataset is an instance of DS 3.1 - State-action space for AI agent interaction with the digital environment.<br><br>The dataset contains structured training, validation, and test data comprising failure and survival events observed in transmission power grid simulations. These were generated using Grid2Op with the WCCI 2022 L2RPN environment. Each data instance is labeled with one of four classes, representing survival or |

impending failure in 1, 3, and 5 timesteps. This dataset was used to train, validate, and test machine learning models that predict grid agent failures in the topology optimization task.

This dataset was developed for and used in the paper titled "Fault Detection for Agents in Power Grid Topology Optimization: A Comprehensive Analysis" by Malte Lehna, Mohamed Hassouna, Dmitry Degtyar, Sven Tomforde, and Christoph Scholz, presented at the Workshop on Machine Learning for Sustainable Power Systems (ML4SPS), part of ECML PKDD 2024. While the paper is pending formal publication, a preprint version is available on arXiv.

**Citation**: Lehna, M., Hassouna, M., Degtyar, D., Tomforde, S., & Scholz, C. (2024). Structured Power Grid Simulation Dataset for Machine Learning: Failure and Survival Events in Grid2Op's L2RPN WCCI 2022 Environment (Version 1) https://doi.org/10.5281/zenodo.13948340

**TABLE 7 – OPENLY AVAILABLE DATASETS**

### 3.1.2 PENDING PUBLIC AVAILABILITY

AI4REALNET identifies several datasets that may be made publicly available, subject to data protection requirements, ethical review, or strategic relevance. These datasets are expected to be valuable for the wider research and innovation community. Table 8 lists the datasets that are considered to be made openly accessible throughout the project lifecycle, for instance, together with the corresponding deliverables or publications.

| Datasets | Release considerations |
|---|---|
| **DS. 1.2 - Datasets for the use cases and digital environments** | Embedded in the digital environments. To be released via software deliverables D1.2-D1.4. |
| **DS. 1.3 - Interactive AI with order-agnostic architectures** | To be released with WP3 software deliverables D3.2 and D3.3. |
| **DS. 2.1 - Experimental and operational data collection for operator personalized biomarkers** | May be shared via INESC TEC open repository after anonymization and curation (end of WP2 or WP4) |
| **DS 2.2 - Socio-technical analysis with dispatcher and traffic controller - Interviews** | Only summarized and anonymized results will be included in Deliverables D2.3 and D3.4. |
| **DS 2.3 - Socio-technical analysis with dispatcher and traffic controller - Observation** | Anonymized summaries will be included in Deliverables D2.3 and D4.3. |
| **DS 2.4 - Socio-technical analysis with dispatcher and traffic controller - Questionnaire** | Aggregated results will be included in Deliverables D2.3 and D4.3. External repository availability depends on |
| **DS. 2.5 - Knowledge-assisted AI performance**<br>**DS. 2.6 - Hierarchical RL in relational domains** | With the publication of the relevant algorithms in WP2 deliverables, namely in software releases of D3.2 and D3.3. |
| **DS. 3.1 - State-action space for AI agent interaction with the digital environment.** | Curated versions to be published with WP3 deliverables. Early examples may be released sooner.<br>**Note:** An instance of this dataset is already openly available (see Table 8). |

| Datasets | Release considerations |
|---|---|
| **DS 4.2 - Test results from the evaluation of AI technical performance** | To be included in the project work reporting in Task 4.3, results. Data is expected to be collected from October to March for the first validation campaign, and from April 2026 to March for the second validation campaign. |
| **DS 4.3 - Data of network operator interactions and feedback produced during Human-AI evaluation scenarios** | To be anonymized and published in Task 4.3 results. Data will be collected during the second validation campaign from April 2026 to March 2027. |
| **DS. 4.4 – Test datasets for AI functional testing.** | Dataset generation code to be released between M30 and M42. |
| **DS. 4.5 - Feedback collected regarding the adoption of AI** | One-time anonymized survey. To be published in Task 4.4 results. Data will be collected during the second validation campaign from April 2026 to March 2027. |
| **DS. 4.6 - Human-in-the-loop Validation Data for AI-Supported Air Traffic Management.** | It will depend on approval from ATCo/planner consent. Data will be collected within the WP4 timeline. Trials are expected to start in Q2-2026. |
| **DS 5.2 - Dataset provided for AI competitions** | It will be generated before the start of the competition as part of Task 5.2, between September 2025 and June 2026. |

**TABLE 8 – DATASET RELEASE CONSIDERATIONS**

### 3.1.3 INTERNAL ACCESS ONLY

The following datasets are not intended for public release, either due to personal or operational sensitivity, ethical/legal restrictions, or because their use is strictly limited to project coordination and internal management.

- **DS 1.1. - Feedback from the participants in the stakeholders´ workshops**

It is not expected that this dataset will be made publicly available, but only disclosed to the meeting´s participants. The main insights were included in Deliverable D1.1 to inform the development of use cases and overall framework design.

- **DS. 4.1 - Test datasets for evaluation of AI technical performance**

This dataset is not expected to be publicly available and will be disclosed only to the meeting participants. It will be created at the beginning of WP4 and may be updated throughout the WP, if necessary. It is expected to be prepared between October 2025 and September 2026.

- **DS 5.1 – Stakeholders' map and contacts**

This dataset contains contact details, organizational affiliations, and feedback collected from ecosystem stakeholders. It is intended exclusively for internal use by the WP5 leader (INESC TEC) to support engagement and dissemination activities. Due to the inclusion of personal information, the dataset will not be published or shared beyond its defined operational purpose.

## 3.2 DATA REPOSITORY SELECTION

The AI4REALNET strategy for the publication of datasets will consider a set of different repositories to identify the repository solution that best fits the data requirements and targeted audiences. Therefore, AI4REALNET will adopt a flexible approach to enable the selection of the most appropriate data repository for each dataset, from generalist repositories to discipline or data-specific ones.

Regardless of their typology, repositories must meet the following conditions:

- Provide a persistent identifier.
- Apply for an open license.
- Ensure that the datasets are openly available with respect to the preservation needs of specific datasets.
- Enable the definition of access conditions if required.
- Are OpenAIRE compliant.

Moreover, AI4REALNET will also comply with existing data policies implemented by the journals where project results will be published, mainly if the submission to a specific repository is a condition for publishing.

In general, the catch-all repository Zenodo, a service hosted by CERN, will be adopted as a data repository for AI4REALNET, as it satisfies the aforementioned conditions if a better-suited repository is not identified for specific datasets. For this purpose, an AI4REALNET community was created on Zenodo from the start of the project. Moreover, Zenodo, as part of the OpenAIRE infrastructures, contributes to the open data movement in Europe.

**Note:** If AI4REALNET partners already have data repository instances in other data repository services, their use is encouraged to streamline data dissemination. However, project outputs deposited outside the Zenodo Community must be reported to the DMP manager so that they are registered in the DMP and added to the datasets in the online version of the DMP with the necessary identifiers.

### 3.2.1 AI4REALNET ZENODO COMMUNITY

The AI4REALNET community policy specifies that it is for the exclusive use of sharing research outputs related to the project. All project partners with a Zenodo account can add their research outputs to the community.

The publication of content added by project partners requires the approval of the community curator, the DMP manager, who in turn will list the available outputs in the DMP updates.

The AI4REALNET Community can be accessed via: https://zenodo.org/communities/ai4realnet

The uploads to the AI4REALNET Community can be made via:
https://zenodo.org/uploads/new?community=ai4realnet

In each upload, AI4REALNET must be added to the funding awards section, as depicted in Figure 1.

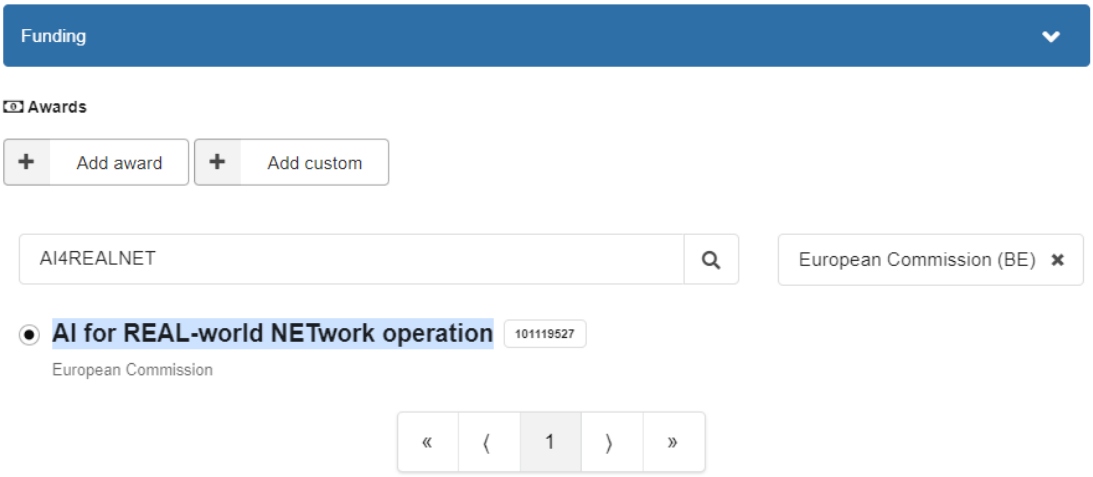


**FIGURE 1 – ADDING AI4REALNET FUNDING AWARD IN ZENODO**

### 3.2.2 GITHUB REPOSITORY SOFTWARE IDENTIFIER

In the end of the project, AI4REALNET GitHub repositories will also be archived in Zenodo so that they are minted with a DOI if these repositories are open access and a license is clearly defined.

To preserve their GitHub repository in Zenodo, AI4REALNET partners can access this page: https://zenodo.org/account/settings/github/

In this process, it must be ensured that the registration is associated with the AI4REALNET Community.

### 3.2.3 INSTITUTIONAL DATA REPOSITORY

The INESC TEC institutional data repository can also support the publication of AI4REALNET datasets. Any user can access the INESC TEC data repository at any time and from any location. Therefore, the deposit data will be available for both project members and the general research communities for as long as the data is deposited in the INESC TEC data repository. The repository is registered in the repository's directory re3data.org.

Deposited datasets are described with a minimal set of Dublin Core metadata, including citation information and the DOI minted via the DataCite Fabrica service.

The deposit of data is the responsibility of the DMP manager and data steward at INESC TEC, and it must be coordinated with the responsible(s) for managing the datasets in each specific task or WP. The data backups are performed daily, while tape backups are performed weekly.

### 3.2.4 AI-ON-DEMAND PLATFORM

As part of the AI4REALNET strategy for data sharing, the project will be registered in the AI-on-demand platform. AI4REALNET will provide use-cases and AI assets (i.e., open-source software developed in WPs 2 and 3, digital environments from WP1) for strengthening the AI-on-demand platform catalog and sustaining the AI4REALNET concept/software beyond the lifetime.

The virtual lab can also host digital environments and make them available to the AI community.

## 3.3 METADATA

Both the INESC TEC data repository and Zenodo already follow appropriate metadata standards, specifically Dublin Core, a highly adopted, domain-agnostic metadata standard published as ISO Standard 15836 :2017. Metadata records based on the Dublin Core standard promote findability and interoperability.

Therefore, a minimal metadata set for each dataset must consider the following categories:

- Administrative metadata for resource management (access rights, license).
- Descriptive metadata to enable the discovery, identification, and selection of datasets (keywords, authors, dates, file size, and format).
- Structural metadata describing hierarchical structures between resources.

Metadata elements for capturing the scientific production context of each dataset should be evaluated on a case-by-case basis. The use of standards for data description and encoding formats will depend on the requirements of each dataset and the setting in which they will be shared.

## 3.4 RECOMMENDED PRACTICES FOR AI4REALNET PARTNERS

### 3.4.1 DATA DOCUMENTATION

Data sharing within and outside the project must include the necessary metadata and documentation for others to understand and reuse the data.

Data documentation includes research protocols, codebooks, software syntax, equipment setting information, and instrument calibration.

AI4REALNET's data may need to be documented at various levels, and a Readme file, in the absence of another type of file, must be included as part of the dataset.

### 3.4.2 LICENSING

AI4REALNET datasets will be made available under Creative Commons (CC) by default. CC licenses are well-suitable for research data due to their conformity to both copyright law and database rights, and at the same time, they are easily readable by end-users.

The license Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) specifies that:

- Credit must be given to the creator.
- Adaptations must be shared under the same terms.

CC Zero is also a license advocated for sharing research data, allowing end-users to use and reuse data without restriction.

In the case of code repositories to be registered in Zenodo (essential to be registered as part of the project products in OpenAIRE), a point of access will be created to the code repository by defining the most suitable relation between the records in *Related Works,* e.g. [*type of relation* – Software*: URL of the source code*], or in the *Software* section by completing the fields related to the Repository URL, Programming Language, and Development Status. Therefore, the relation to the original resource will

be clearly defined in the *External Resource* section in Zenodo [Available in], ensuring that users will have access to licensing information about the code repositories.

### 3.4.3 FILE NAMING

File names are an essential data identifier throughout the data life cycle. Not only does it make it easy to access data, but it also helps to understand what a data file is and its content. File names must remain meaningful and valuable beyond their original creation and storage location, particularly in shared environments. To allow ease of access and understanding of the data, some elements must be considered when implementing the file name strategy. The file name must provide metadata about its content.

Recognizing that the same strategy cannot be applied to the diversity of datasets that AI4REALNET will generate, the DMP does not define a policy for naming datasets, enabling the necessary flexibility for the data creators to specify the most suitable convention for their folders and file's structure. However, when adopting a file name strategy, elaborating a quick access guide is recommended, enabling partners to decode the elements that make up the filename quickly.

Despite the necessary flexibility, some conditions must be considered:

- Context information. The components of the file names must include specific or descriptive information (essential attributes).
- Consistency. The adopted convention must be followed systematically.
- Short and relevant names.
- The use of special characters is to be avoided.
- The use of underscores or hyphens is recommended over full stops or spaces.

### 3.4.4 FILE DATE FORMATTING

Dates included in the file name must follow the format Year-Month-Day to maintain the chronological order and simplify the process of sorting and browsing data files. This ensures compliance with the international standard for the representation of date and time, ISO 8601 (https://www.iso.org/iso-8601-date-and-time-format.html).

### 3.4.5 VERSIONING

For versioning purposes, it is common practice to use consecutive numbering for major version changes, with decimals used for minor changes (v1; v1.1; v2.1; v2.2). However, too many similar or related files can make it confusing to access the data as the number of files grows. The priority is to retain a copy, if possible, of the original "raw" data or definitive version, together with a well-documented record of changes.

### 3.4.6 DATA AVAILABILITY STATEMENT

The Data Availability Statement may be required for the manuscript submission workflow. When publishing project results, AI4REALNET partners will consider the following elements in their Data Availability Statement:

- **For open datasets:** data repository or service from which the dataset can be accessed, DOI, citation information, license, list of data items.

- **For datasets with conditions of access:** the ethical reason why the dataset cannot be shared openly, link to a permanent record detailing restrictions and conditions for access.
- **Third-party datasets:** citation information.

# 4. DATA SECURITY

This section outlines a preliminary set of recommendations regarding data storage and backup, to be refined locally, depending on each partner's infrastructure. AI4REALNET also recommends particular attention to the requirements stated in ISO 27001 for the purposes of information security management, which can be further detailed in the following updates of this DMP.

To prevent the risk of data loss, it is recommended that data be stored in institutional network drives, which must be routinely backed up, and account authentication systems must be provided to prevent unauthorized access, preferably by resorting to strong passwords. Moreover, if the data contains sensitive or personal information, the use of VPN is recommended.

To manage and share data securely, specific data storage platforms are not recommended for long-term storage:

- External portable storage devices, such as external hard drives and USB drives, given their longevity uncertainty and how easily these can be damaged or lost.
- Popular cloud storage platforms aimed at the general public.

Personal Computers and Laptops are very useful for daily work. However, the DMP will not specify or recommend how partners should manage their data in this context. However, project partners must ensure that data backup is regularly made through suitable networked drives.

There is no one-size-fits-all approach, and AI4REALNET acknowledges the necessary autonomy for project partners when adopting the most suitable approach to data security.

- Data must be stored in at least two different media (if possible, three).
- Daily backup of network drives as primary data storage device.
- Weekly backup for long-term off-site data preservation.

## 4.1 STORAGE AND BACKUP

Most datasets in AI4REALNET will be stored in the INESC TEC Drive**,** a secure on-premise file access and sync platform based on Nextcloud. This infrastructure supports daily automated backups and offline storage procedures. It is hosted on INESC TEC's secure servers and integrates with the institutional Lightweight Directory Access Protocol (LDAP) system, which ensures that accounts are centrally managed and password reset policies are enforced without intervention from IT staff. A detailed description of the platform and security protocols is provided in Annex 3.

Depending on the nature of the dataset and the responsible partner, alternative or complementary storage solutions may be used, such as Surfdrive cloud storage, local institutional servers.

Table 9 provides an overview of the storage and backup strategy for AI4REALNET datasets.

| Dataset | Storage and backup strategy |
|---|---|
| • **DS. 1.1 - Feedback from participants in the stakeholders' workshops**<br>• **DS. 2.1 - Experimental and operational data collection for operator personalized biomarkers**<br>• **DS. 4.3 - Test datasets for AI functional testing** | These datasets will be stored in the INESC TEC drive with daily backups and offline storage. |

| Dataset | Storage and backup strategy |
|---|---|
| • **DS. 4.4 - Data of network operator interactions and feedback produced during Human-AI evaluation scenarios**<br>• **DS. 4.5 - Feedback collected regarding the adoption of AI**<br>• **DS. 4.6 - Validation data of air traffic control**<br>• **DS. 5.1 - Stakeholders map and contacts**<br>• **DS 5.2 - Dataset provided for AI competitions** | |
| **DS. 3.1 - State-action space for AI agent interaction with the digital environment** | The files can be stored in INESC TEC drive; however, individual instances that may differ slightly will be held independently (and locally) by each partner using the digital environment. |
| • **DS. 2.2, DS. 2.3, DS. 2.4 - Socio-technical analysis with dispatcher and traffic controller**<br>• **DS. 4.1 - Test datasets for evaluation of AI technical performance**<br>• **DS. 4.2 - Test results from the evaluation of AI technical performance** | Local PC and INESC TEC drive. |
| • **DS. 2.5 - Knowledge-assisted AI performance**<br>• **DS. 2.6 - Hierarchical RL in relational domains**<br>• **DS. 1.3 - Interactive AI with order-agnostic architectures** | Files are stored locally, and backups are created using Surfdrive cloud storage. Storage on safe and backed-up cloud storage (e.g., Surfdrive) until publication in a reliable repository (e.g., UvA instance of Figshare). |
| **DS. 1.2 - Datasets for the use cases and digital environments** | The files can be stored in the INESC TEC drive. When curated, the project's GitHub will include dataset and data generation functions. |

**TABLE 9 – STORAGE AND BACKUP**

## 4.2 PRESERVATION

Preservation of data in AI4REALNET is guided by ethical, legal, and scientific principles. Datasets that do not contain personal or sensitive information and are suitable for open sharing will be preserved for long-term reuse in trusted repositories, in accordance with FAIR principles and repository policies. For datasets involving personal or sensitive data, retention will be strictly limited to what is necessary to fulfil the research purposes.

As a general reference, such data may be retained for up to three years after the project ends, if needed to support scientific validation, reporting, or audit processes. However, partners are expected to review personal data regularly and to delete or irreversibly anonymize any data that is no longer required, in line with the GDPR obligations and the project's ethical commitments.

Table 10 outlines specific preservation needs and exceptions identified so far.

| Dataset | Preservation needs |
|---|---|
| **DS. 1.1 - Feedback from participants in the stakeholders' workshops** | There is no specific need since this dataset is directly used to elaborate on the use cases and deliverable D1.1. |
| **DS. 2.1 - Experimental and operational data collection for operational personalized biomarkers** | The data collected is intended to be fully anonymized after six months, at this point, all information that could directly or indirectly identify participants will be destroyed. |
| **DS. 4.1 - Test datasets for evaluation of AI technical performance**<br>**DS. 4.2 - Test results from the evaluation of AI technical performance** | The data does not need to be preserved after the project duration since the reuse value is low after that period. What is essential to preserve is the code that conducts the functional tests and ensures the reproducibility of the results. |
| **DS. 2.2, DS. 2.3, D. 2.4 - Socio-technical analysis with dispatcher and traffic controller** | No need to preserve full audio recordings. The processed outputs (e.g., models, requirements) are sufficient to preserve.<br>Full raw observation data does not need to be preserved. Processed models may be preserved if relevant.<br>Raw data may be discarded post-analysis. Preservation of key analytical outputs is advised. |
| **DS. 4.4 - Data of network operator interactions and feedback produced during Human-AI evaluation scenarios** | There is no need for preservation. The output of data processing (such as knowledge mapping, functional resonance model, etc.) may need to be preserved. |
| **DS. 2.5 - Knowledge-assisted AI performance**<br>**DS. 2.6 - Hierarchical RL in relational domains** | Data can, in principle, be re-generated as it is from a synthetic environment. Preservation is desirable but not critical. |
| **DS. 1.3 - Interactive AI with order-agnostic architectures** | Possible human interaction is the most crucial part to be preserved. It cannot be re-generated and offers the possibility of being reused (e.g., to train and evaluate other algorithms). |
| **DS. 5.2 - Dataset provided for AI competition.** | The raw data should be preserved for 12 months. The consolidated results should be preserved for at least two years after the project. The code in GitHub will ensure the reproducibility of the test scenarios. |

**TABLE 10 – PRESERVATION NEEDS**

# 5. LEGAL COMPLIANCE

AI4REALNET ensures that all activities involving personal data or human participants are carried out in full compliance with applicable legal and regulatory requirements. The project embraces a data protection-by-design and by-default approach and upholds the fundamental rights of data subjects as defined under the GDPR.

Partners are expected to manage data responsibly and to follow their internal procedures for ethical review, including seeking ethics committee approval before collecting personal data or data from human participants.

Informed consent is required for all activities involving human participants, but it is recognized as part of a broader ethical assessment. Depending on data sensitivity, risk level, and institutional frameworks, ethical approval may not be mandatory, provided this is confirmed by internal procedures. No data collection involving personal or sensitive information will proceed without prior approval by the responsible institutional body.

While AI4REALNET defines a set of general legal and ethical principles applicable to all activities involving personal data and human participants, each dataset may require specific safeguards or procedures depending on its nature, context, and sensitivity.

Table 11 outlines key ethical considerations for datasets identified as involving personal data collection or human participation, including planned procedures for ethical approval, informed consent, anonymization, and data minimization. This information reflects the current state of planning and implementation and will be updated throughout the project lifecycle as protocols are finalized or adjusted.

| Data Collection and Ethical Considerations |
|---|
| **DS. 1.1 - Feedback from participants in the stakeholders' workshops**<br>Only the name and feedback (as professional feedback on the use case presented) have been collected. |
| **DS. 2.1 - Experimental and operational data collection for operational personalized biomarkers**<br>Only anonymous and aggregated results may be disseminated or published in scientific publications (which may involve research teams from different institutions) with the consent of the participants (information available on the informed consent document).<br>This dataset will mainly contain physiological data such as electrocardiogram, respiration, and core temperature estimation, collected through a wearable chest band (VitalSticker). It also contains psychological and contextual data, including sociodemographic information, and stress and fatigue levels self-reported via questionnaires.<br>Before the study, a description of the project and study objectives, jointly with a protocol and informed consent, will be prepared and submitted to the ethics commission and the DPO of INESC TEC. The study will only move forward after positive feedback from the ethical committee and with the guidance and monitoring of the DPO. After this feedback, all the experiments will be prepared according to the defined protocol. The data collection process will start only if the operator accepts free will to participate in the study and signs the informed consent. Initially, data will be pseudorandomized, and a correlation between the data and the human subject will only be kept by the project partners who employ the participants. The pseudonymization process and all the data curation procedures, and even possible data erasing, are defined in the informed consent and will be explained to the participant. The study is estimated to run in 2026. The estimated duration is still not defined due to a project undergoing tasks that need to be further defined. |
| **DS. 1.3 - Interactive AI with order-agnostic architectures** |

If a dataset is extended to include human interaction data, standards for anonymization and privacy will be followed. Only decisions made will be logged; this explicitly excludes personally identifiable information.

**DS. 2.2; DS 2.3; DS 2.4 - Socio-technical analysis with dispatcher and traffic controller**

- **Interviews**
- **Observation**
- **Questionnaire**

The analysis focuses on the daily tasks performed by individuals, breaking them down into subtasks to distinguish between those carried out by humans and those handled by technology. It investigates the decision-making process—who is responsible, how decisions are made, and who is involved—alongside the knowledge required for each subtask and how it is acquired. Special attention is given to subtle indicators such as coordination breakdowns, information gaps, and verbal cues. In the context of observations, the same analytical approach is applied, emphasizing real-world task execution to extract socio-technical insights.

At FHNW, an internal data management task force has created guidelines and documents for data collection. These guidelines were followed and explained to the interviewers. Interviewees had to sign a document confirming that they understood the data policy and agreed to these conditions. These informed consents are stored only on secure FHNW servers.

The audio files from the interview are only stored on our FHNW server in Switzerland. The audio file's transcript does not contain any personal data, and identifying characteristics are replaced by a pseudonym (e.g., a number). Only the transcript is shared within the AI4REALNET community, and no other information is distributed. The transcript of the observation does not contain any personal data, and identifying characteristics are replaced by a pseudonym (e.g., a number). No other information is distributed within the ai4realnet community.

- **DS. 4.3 - Data of network operator interactions and feedback produced during Human-AI evaluation scenarios**
- **DS. 4.5 - Feedback collected regarding the adoption of AI**

Both datasets are planned to include subjective and physiological information collected from participants involved in Human-AI interaction and evaluation activities. The data to be gathered will cover participants' professional roles and responses to structured questionnaires on sociodemographic characteristics, perceptions and opinions about AI and its implications in professional contexts, stress and fatigue levels, and agreement ratings on a 0–100 scale. A broad set of validated Likert-scale instruments will also be applied, addressing topics such as trust in AI, explainability, usability, workplace dynamics, human-AI collaboration, and related psychological constructs (e.g., XAI Survey, KIAM, SAGAT, WDQ, TSST, among others). Biometric signals such as ECG, heart rate, respiration, estimated core temperature, posture, and activity will be collected via a wearable chest band with textile electrodes. Additionally, a Visual Analogue Scale will be used to assess stress and fatigue.

Prior to data collection, the study will be submitted for approval by the relevant ethics committee, ensuring compliance with institutional ethical procedures.

All participating institutions are committed to confidentiality and to the non-disclosure of personal data. The data will be processed in a pseudo-anonymous and aggregated manner, ensuring that individual identities are not revealed. Several measures will be implemented to protect participants' privacy:

A) Each participant will be assigned a unique code.
B) All data will be stored securely on local INESC TEC servers with access strictly limited to the research team.
C) After six months, the datasets are intended to be fully anonymized, with the destruction of all information that could directly or indirectly identify participants.

| |
|---|
| D) Any video recordings containing identifiable visual data will be stored securely for the duration of the six-month project and will be accessible only to authorized researchers. |
| **DS. 4.6- Human-in-the-loop Validation Data for AI-Supported Air Traffic Management**<br><br>The specific legal and ethical procedures applicable to this dataset are yet to be defined. However, all data collection and processing activities will be evaluated in accordance with the general ethical and data protection principles of the AI4REALNET project. Only essential personal data will be collected, retained no longer than necessary, and processed in compliance with GDPR. Appropriate safeguards such as anonymization or pseudonymization will be applied wherever relevant. Ethical approval and informed consent will be required for any activity involving human participants, ensuring voluntary participation and the right to withdraw at any time. |
| **DS. 5.2 - Dataset provided for AI competition.**<br><br>Possible ethical or legal issues will be managed before publication in the context of AI competitions. |

**TABLE 11 – EXPECTED ETHICAL OR LEGAL ISSUES**

While Table 11 provides a detailed overview of the ethical and legal procedures applicable to each dataset, the following Data Flow Matrix (Table 12) outlines how data are expected to be transferred between partners, indicating sources, transfer conditions, and safeguards. This information is a preliminary assessment and will be further refined as protocols are consolidated.

| Dataset | Data Flow Matrix |
|---|---|
| **DS. 2.1 - Experimental and operational data collection for operational personalized biomarkers** | INESC TEC → anonymized/aggregated results → AI4REALNET Partners<br><br>**Source:** INESC TEC<br><br>**Sharing Conditions:** Only summary results are shared with partners (raw data remain at INESC TEC; any link to individual participants is kept only by their employers).<br><br>**Safeguards:** Informed consent; ethics approval; pseudonymization; restricted access. |
| **DS. 2.2; DS 2.3; DS 2.4 - Socio-technical analysis with dispatcher and traffic controller**<br>• **Interviews**<br>• **Observation**<br>• **Questionnaire** | FHNW → pseudonymized transcripts/aggregated survey results → AI4REALNET Partners<br><br>**Source:** FHNW<br><br>**Sharing Conditions:** Transcripts and survey results shared in pseudonymized or aggregated form; original audio files and consent forms remain at FHNW<br><br>**Safeguards:** Informed consent; pseudonymization; secure storage on FHNW servers; restricted access. |
| **DS. 4.3 - Data of network operator interactions and feedback produced during Human-AI evaluation scenarios** | RTE→ pseudonymized multimodal data → AI4REALNET Partners<br><br>**Source:** RTE |

| | |
|---|---|
| | **Sharing Conditions:** Multimodal data (e.g. ECG, questionnaires, video) shared in pseudonymized or aggregated form; no identifiers are shared.<br>**Safeguards:** Ethics approval; informed consent; restricted access; secure storage.<br>**Retention:** Identifiers are destroyed after six months; datasets intended to be fully anonymized thereafter. |
| **DS. 4.5 - Feedback collected regarding the adoption of AI** | RTE→ pseudonymized multimodal data → AI4REALNET Partners<br>**Source:** RTE<br>**Sharing Conditions:** Feedback from questionnaires and interviews shared in anonymized or aggregated form; no personal identifiers included.<br>**Safeguards:** Informed consent; minimization of sensitive data; pseudonymization where applicable; restricted access.<br>**Retention:** Identifiers are destroyed after six months; datasets fully anonymized thereafter. |
| **DS. 4.6 - Human-in-the-loop Validation Data for AI-Supported Air Traffic Management** | TU DELFT →to be defined → AI4REALNET Partners<br>**Source:** TU DELFT<br>**Sharing Conditions:** To be defined once validation protocols and ethics approvals are in place; only essential data will be shared.<br>**Safeguards:** GDPR compliance; Informed consent; ethics approval. |

**TABLE 12– DATA FLOW MATRIX**

## 5.1 DATA MINIMIZATION AND STORAGE LIMITATION PRINCIPLES

AI4REALNET must only collect essential personal information data, which shall be retained no longer than necessary. This means that personal data should only be kept for the time necessary to carry out the purposes for which it was collected or the time required to comply with applicable law.

Personal data must be reviewed periodically to decide whether unnecessary identifying information is retained. Data retention limits are set until the end of the period needed to conduct the research.

Personal information must be safely deleted or destroyed if it is no longer needed. AI4REALNET will consider using appropriate software for the deletion of data and encryption key deletion to prevent the recovery of data stored in an encrypted container or disks.

Specific data protection procedures for destroying data must be defined in compliance with GDPR and the applicable legal framework.

## 5.2 APPROPRIATE SAFEGUARDS: ANONYMIZATION AND PSEUDONYMISATION

Personal data processing for scientific research purposes or statistical purposes shall be subject to appropriate safeguards for the rights and freedoms of the data subjects.

These safeguards include technical and organizational measures such as pseudonymization and data anonymization procedures, which must be evaluated and carried out by project partners, while considering the realization of the project objectives. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner. Therefore, for instance, when the identifying information is no longer needed, direct identifiers should be removed, where possible, by deleting them or replacing them with pseudonyms.

## 5.3 VOLUNTARY PARTICIPATION AND INFORMED CONSENT

AI4REALNET partners shall implement ethical procedures to ensure the voluntary participation of human subjects in project activities/ studies, as well as the respect for the fundamental ethical principles of autonomy, beneficence, non-maleficence, and justice, and applicable data protection legal requirements. The partners involved in activities or studies involving human subjects or personal data shall comply with their internal procedures for ethical assessment and clearance (namely, the approval of an institutional review board/ethics committee) before those activities take place. As a default rule, the leading partner in such an activity or study will be in charge of getting ethical approval.

AI4REALNET partners will always ask for informed consent in all activities involving human participants, provided the approval of an ethical committee, explicitly stating that participation is voluntary and that anyone has the right to refuse to participate and to withdraw their participation or data at any time without consequences. Annex 4 provides an illustrative, provisional example concerning a specific study to be led by the Coordinator, INESC TEC.

Human participants will be informed about the aims, methods, and objectives of data collection, the benefits, and risks, as well as in which conditions the data may be shared, in accordance with the requirements stated in Article 13º or 14º of GDPR.

Notwithstanding the above, each partner shall assess the legal basis applicable to the respective activities/ studies whenever personal data processing is required and perform in a timely manner the assessment as to the need to perform a DPIA. Furthermore, we recall, quoting the EC guidelines on identifying ethics issues in EU-funded research, that: *"Even if service providers and external collaborators are engaged in the research, the obligation to safeguard data subject's rights and freedoms rests with the principal researchers (e.g., the beneficiary and partners of a consortium). This obligation cannot be 'outsourced' or delegated (e.g., when surveys are conducted or data is processed or hosted by third parties or subcontractors).".*

## 5.4 ROLES AND RESPONSIBILITIES

Partners involved in personal data processing activities shall assess their roles and responsibilities in relation to such activities according to the GDPR (as data controllers, joint controllers, or data processors) and, whenever necessary, shall establish the adequate contractual instruments in

accordance with the law and the consortium agreement. Also, internally, project **partners shall take care of an adequate attribution of roles and responsibilities among teams involved in data processing and put in place suitable policies, procedures, and organizational measures, including continued** training, in order to ensure compliance.

Project partners collecting personal data must have a person responsible for monitoring GDPR compliance. The DPO, when applicable.

## 5.5 DOCUMENTATION

Partners who process personal data maintain a record of processing activities in accordance with Article 30 of the GDPR.

# 6. ETHICS COMPLIANCE ASSESSEMENT

While ethical considerations and data governance are central to the responsible development of AI, it is important to clarify that we are not conducting formal ethical or legal compliance assessments as defined under the AI Act. Such assessments are typically conducted over pre-commercial AI-based products and within the responsibility of independent ethical bodies and certified legal entities. Instead, our approach focuses on conducting Trustworthy AI assessments in accordance with the European Commission's framework for Trustworthy AI[1], which forms the basis for the AI Act and the Assessment List for Trustworthy Artificial Intelligence (ALTAI). This continuous assessment process enables us to address key trustworthiness dimensions, such as transparency, robustness, privacy, and accountability, throughout the lifecycle of AI solutions developed within the project.

By embedding these assessments into our methodology, we also aim to promote alignment of the project's activities with the AI Act's objectives and facilitate the future regulatory compliance of AI systems derived from the AI4REALNET outcomes. This proactive perspective ensures that trustworthiness is integrated by design, laying the groundwork for more formal assessments where necessary in subsequent productization phases. By doing so, it improves readiness for AI-act compliance of future solutions based on AI4REALNET outcomes. This section outlines the progress achieved by the consortium in ethics compliance assessment, guided by an integrated trustworthiness-by-design approach and supported by a multidisciplinary team. References to relevant project deliverables and activities are provided to facilitate a comprehensive analysis of the project's methodology and outcomes.

## 6.1 ALTAI ANALYSIS IN USE CASES DESIGN PHASE

In WP1, in particular for Task 1.1 during the use cases formal writing and requirements analysis, the ALTAI assessment tool was used to identify the relevant risks and ethical concerns and translate them to non-functional (and functional) requirements in the use cases, in alignment with the framework for trustworthy AI established by the high-level expert group on artificial intelligence appointed by the European Commission. This process also allowed us to evaluate the suitability of applying ALTAI at the early stages of development, identify limitations, and provide recommendations for its improvement. These recommendations were included in deliverable D1.1 (Bessa et al., 2024).

The project used the ALTAI tool to perform an *ex-ante* assessment of the use cases in accordance with the framework for trustworthy AI from the European Commission, as depicted in Figure 2. This allowed the consortium to a) identify risks and ethical issues particularly relevant to the considered use cases, b) define use case requirements to be fulfilled by the solutions developed in the project, and c) develop suitable metrics to validate that these requirements are appropriate and sufficient to mitigate the identified risks and ethical concerns.

[1] For simplicity we use this expression, referring to the guidelines and framework produced by the Independent HLEG on AI, set up by the EC, conceptualizing the notion of Trustworthy AI, and at the origin of the regulatory efforts that later resulted in the AI Act

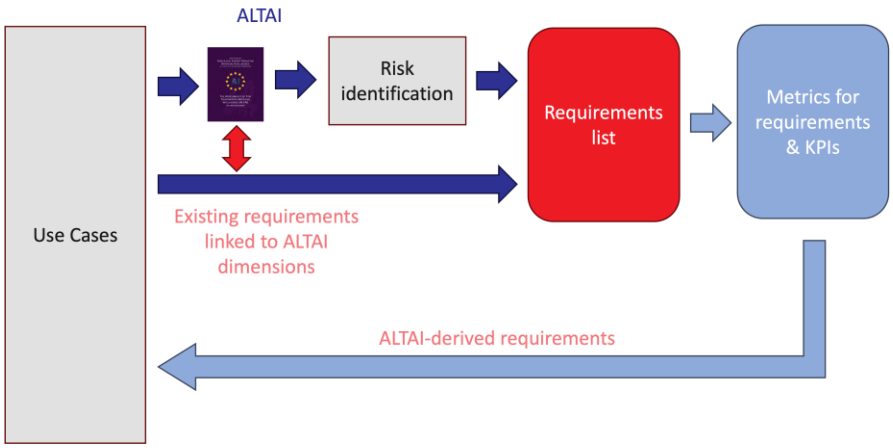


**FIGURE 2 – PROCESS FOLLOWED BY AI4REALNET TO DERIVE NON-FUNCTIONAL REQUIREMENTS FROM ALTAI**

Ethical concerns in AI use cases were identified through domain expert knowledge and internal workshops with consortium partners. These sessions introduced the ALTAI framework and assessed initial versions of each use case to connect stakeholder interests, often reflective of individual or societal values, to the ALTAI dimensions. Following the workshops, participants collaborated via a shared ALTAI questionnaire, contributing insights and annotations. A designated expert oversaw this process to ensure consistency. Each ALTAI question led to a decision: whether the issue was relevant to the use case, along with justifications and corresponding requirements. These decisions were documented in a structured format with columns for the question, decision ("Relevant" or "Not Relevant"), ethical considerations, and linked requirements. This structured assessment supports transparency and ethical justification for the AI4REALNET project and is publicly accessible via its website[2], with results detailed across three application domains, modeled after the format used by (Stefani et al., 2023).

Since each application domain has its own specific characteristics, individual assessments were pursued for the power grid, railway, and air traffic management domains. The detailed outcomes of these evaluations are documented in Deliverable D1.1 (Bessa et al., 2024). A summary of the key conclusions for each domain is presented below. Moreover, for the sake of completeness, Annex 5 presents a summary of the project's use cases.

The fulfillment of ALTAI-derived requirements will be assessed both quantitatively and qualitatively as part of the evaluation activities in WP4 (see more details in Sections 6.2–6.4).

Power grid

The ALTAI assessment of UC in the power grid domain showed the relevance of over 80% for 5 of 7 ALTAI requirements (see Figure 3): accountability, human agency, and oversight, transparency, technical robustness and safety, diversity, non-discrimination, and fairness.

[2] A summary of the answers to the ALTAI questionnaire summary can be found in: https://ai4realnet.eu/wp-content/uploads/2024/08/D1.1-ALTAI_Summary.pdf

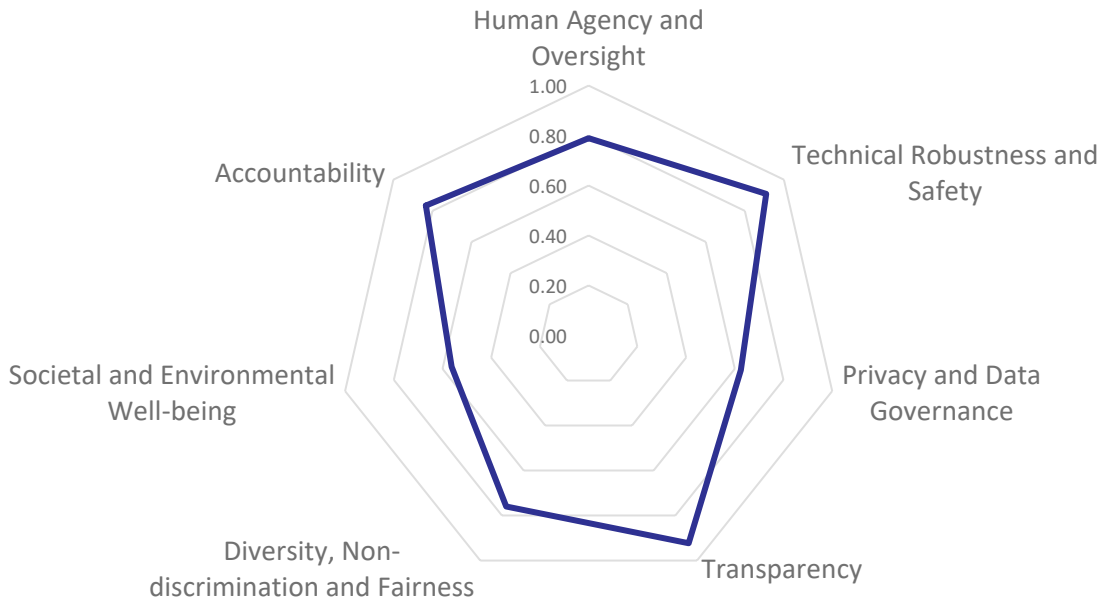


**FIGURE 3 – POWER GRID: RELEVANT ALTAI REQUIREMENTS**

AI systems for power grid operations are designed to enhance human agency while ensuring technical, ethical, and environmental integrity. Operators retain full decision-making control, supported by AI recommendations with alarms in place when AI cannot provide guidance. Training ensures operators understand AI logic, such as reinforcement learning, and can simulate outcomes. Technical robustness is achieved through cybersecurity measures, resilience to data attacks, and performance monitoring via stress tests and robustness metrics. Safety is ensured by maintaining human oversight over AI outputs, minimizing the impact of model inaccuracies.

Privacy risks are minimal as the AI system processes anonymized data and complies with GDPR. Transparency is supported through historical logging of AI decisions and explainability techniques like sensitivity analysis. Operators are alerted to potential AI failures and trained for effective interaction. Fairness is promoted by avoiding bias in AI decisions, particularly in energy dispatch, and involving stakeholders in design and evaluation. Societal and environmental goals are prioritized through carbon-conscious actions and enhanced grid resilience to extreme weather. AI augments but does not replace operators. Finally, accountability is ensured through auditability mechanisms, data traceability, and risk management strategies in compliance with upcoming AI regulations, ensuring system safety and reliability during operational phases.

Railway network

For the railway network, a large proportion of questions in the questionnaire were considered relevant for the UCs but out of scope for the proof of concept (POC) that will be implemented during the AI4REALNET project. The POC is limited to testing in the simulation environments and is concentrated on the technical feasibility of the functional requirements. Hence, many ethical dimensions will not be included for the first implementation due to the use in a controlled environment but are relevant at later stages. Figure 4 shows the relevant ALTAI requirements and plans for implementation in the AI4REALNET project (Railway PoC). The ALTAI questionnaire on the PoC yields requirements on Human Agency and Oversight, Social and Environmental Well-Being, and Transparency.

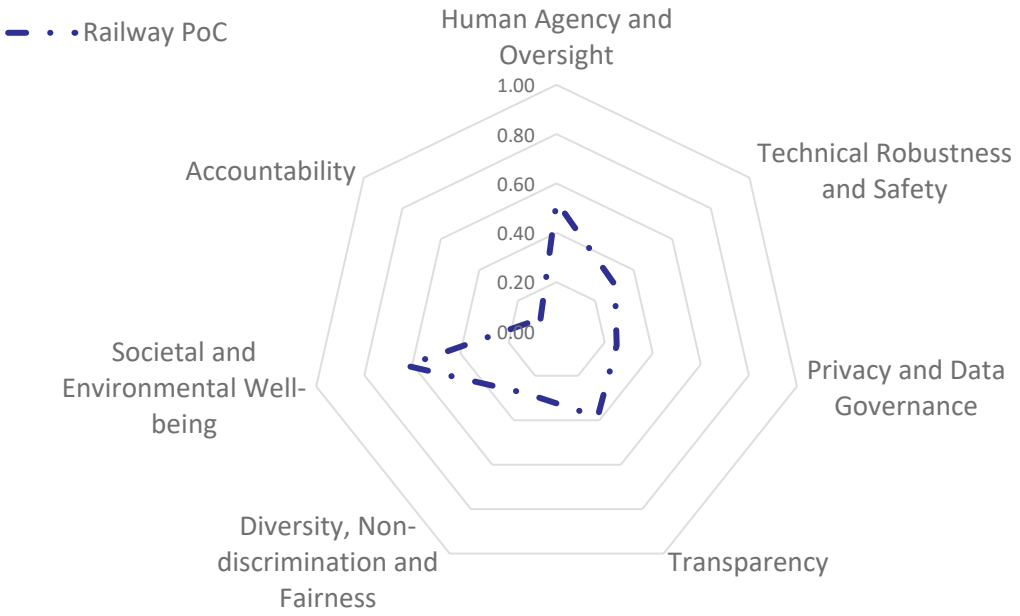


**FIGURE 4 – RAILWAY: ALTAI REQUIREMENTS RELEVANT FOR POC PLANNED FOR AI4REALNET**

Additionally, we performed the ALTAI analysis to identify relevant non-functional requirements for the system's future real-world application. Figure 5 depicts how the number of identified relevant ethical dimensions for the system planned for the application in real-world scenarios increases in comparison to those of the PoC. The dashed line is equivalent to Figure 4 (Railway PoC). The solid line shows the proportion of relevant questions for the real-world applications, to assess the overall coverage of use cases by ALTAI questionnaires. The difference between the dotted line (PoC) and the full line (Full Railway Use Case) illustrates how some ethical requirements become relevant at later stages of development than the ones covered within the AI4REALNET scope. For the complete coverage of the use cases, requirements such as accountability, technical robustness and safety, privacy and data governance, and transparency have grown in importance.

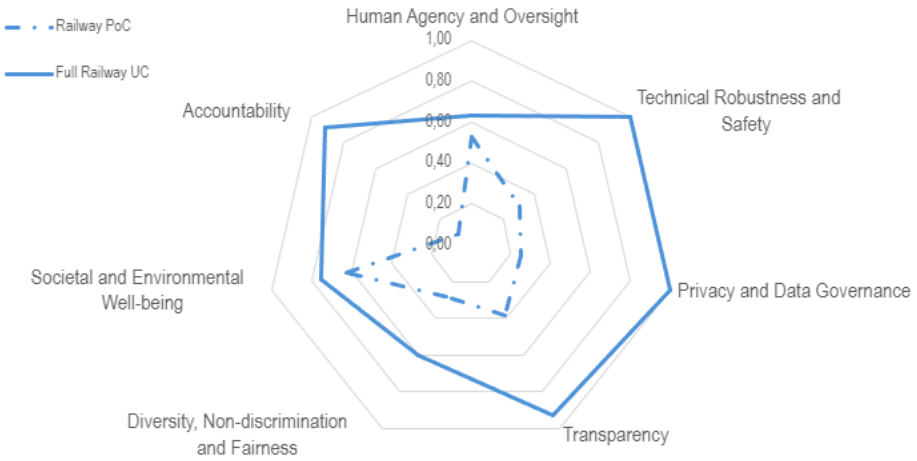


**FIGURE 5 – RAILWAY: ALTAI REQUIREMENTS RELEVANT FOR POC AND EXTENDED VERSION FOR THE REAL-LIFE APPLICATION**

The AI system supports human decision-making in railway operations, with oversight ranging from Human-in-the-Loop to Human-in-Command. Training ensures proper usage and mitigates overreliance risks. Procedures enable safe return of control to humans, with mechanisms to detect adverse effects and regulate self-learning. Technical robustness includes performance monitoring and resilience to attacks, though long-term certification and safety procedures are beyond current scope. Accuracy and human oversight remain critical. Privacy risks are minimal as no personal data is used, and GDPR compliance is not necessary at this stage. Transparency is ensured through traceable logs, explainable outputs, and clear AI-human communication. Fairness is supported by avoiding bias in delay distribution, with stakeholder input guiding development. Societal impact includes increased efficiency and changes to workforce skills, though training development lies outside the current scope. Accountability is upheld through documentation and logging, enabling future auditing and internal ethics monitoring despite the absence of formal risk management at this phase.

Air traffic management

The summary of the ALTAI questionnaire filled for the air traffic management use cases is in Figure 6. The requirement of transparency stands out clearly from the others since in this safety-critical, human-in-the-loop domain, effective collaboration centers on traceability, explainability, and clear communication, enabling operators to understand AI decisions, maintain situational awareness, and uphold trust under a 'management by exception' regime. The focus on its constituents, traceability, explainability, and communication is shaped by the type of AI system described in use cases. For an AI assistant, transparency describes different aspects of the human-AI collaboration and can be used to facilitate the operator's successful use of AI system predictions. The productive cooperation between an operator and an AI system is based on reliable, understandable, and sufficient communication.

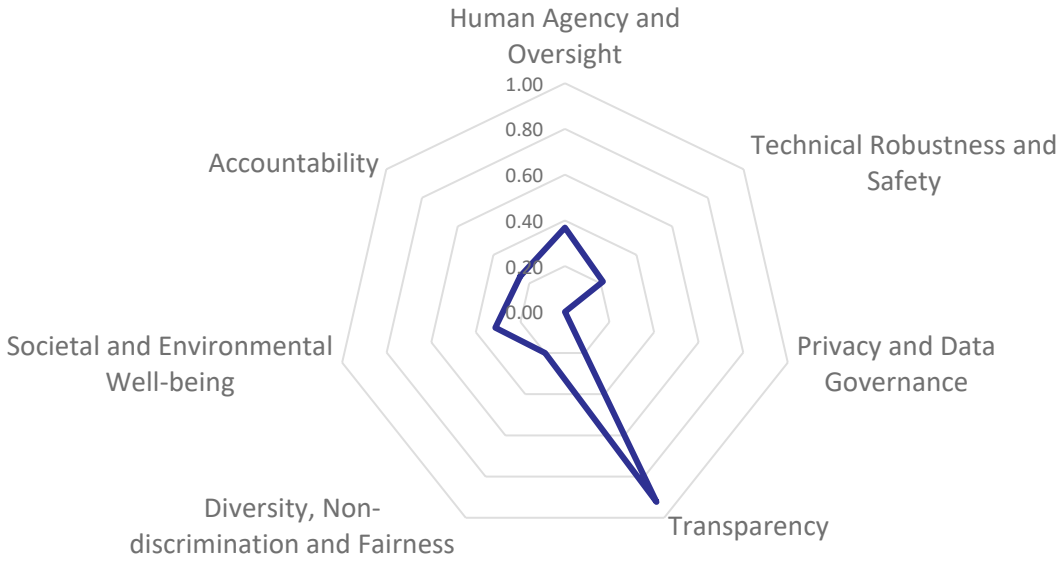


**FIGURE 6 – ATM: RELEVANT ALTAI REQUIREMENTS**

The AI system acts as a recommender, keeping human operators in control, though increasing autonomy may reduce vigilance and shift oversight to "management by exception". Alarms alert operators when AI fails or environmental conditions change. Safety concerns grow with higher automation; resilience to attacks, performance monitoring, and fallback mechanisms are essential. AI

updates, especially through online reinforcement learning, must be logged and communicated to prevent confusion. Privacy risks are minimal, though anonymization is required when personal data from air traffic controller performance and decisions influences KPI calculations. Transparency is supported through traceability of AI decisions, explanation-on-demand, and metrics assessing operator trust. Surveys will monitor human-AI interaction. While fairness issues are not expected, changes in workloads and required skills warrant inclusive design and operator feedback. The system supports societal and environmental goals by improving efficiency and reducing emissions, contingent on effective operator training. Accountability is ensured via detailed logging of model parameters and states, with audits needed after updates to maintain traceability and reliability.

## 6.2 RISK ASSESSMENT

The European Union's Artificial Intelligence Act (EU AI Act; Art 9) demands that a “risk management system shall be established, implemented, documented and maintained” for high-risk AI systems. This requirement underscores the necessity for continuous monitoring and iterative updates to address emerging risks effectively.

Effective AI risk assessment for AI-supported decision making requires a holistic approach that considers the interplay between technological components and social actors. For this reason, in the framework of WPs 1 (Task 1.3) and 4 (Task 4.2 and 4.3), AI4REALNET developed a risk assessment framework that builds on established standards and adopts an epistemological definition of AI-related risks. The framework is based on a multi-component definition of risk traditionally employed in natural disaster risk management. According to this, we assess risks based on four components: Likelihood, Magnitude, Vulnerability, and Exposure. The decomposition of these factors enables the identification of diverse potential impacts on individuals, infrastructure, and both tangible and intangible assets. Moreover, the framework has been adapted to be compliant with ISO (International Organization for Standardization) standards on Risk management, ISO 31000 and ISO/IEC 23894, ensuring a structured and standardised approach to risk evaluation.

An early version of this framework was applied to the analysis of risks related to the “Safety and Robustness of AI Solutions” (Task 4.2), and the framework description and first application to the three sections can be found in Section 4.2 and Annex 3 of Deliverable D4.1 (Lemetayer et al., 2025). For the risk assessment, the multi-component framework was structured in a shared spreadsheet and sent to the industry partners, who filled it according to their understanding of risks in their use cases.

The risks identified in the ATM use case considered technical robustness against cyberattacks, bias against specific routes, input data quality – corrupted or missing information, and the quality of AI performance.

Railway use cases defined risks related to unusual behaviour in data, which were missing during the training phase. The algorithms can fail to adjust to the data shift and new patterns in data, e.g., due to the new traffic patterns or passenger demands. Additionally, extreme or rare weather conditions may fail in the training data and in the simulated environment, and lead to poor performance in real-life conditions, if they appear. Another type of risk was related to the new skills and tasks that the human operator will have to work with after the integration of the decision-making algorithms.

The use cases of the power grid domain also underlined the risks coming from the limited training data, e.g., lack of scalability to real data. Furthermore, this use case also identified the risk to the societal well-being of its primary users with the deskilling of human operators. Additional risks are

coming from the interaction of operators with the systems, e.g., poor interface and low usability were noted as potential risks leading to misinterpretations of the system's decisions by human operators.

We will continue the application of the risk management approach on Tasks 4.3 "Social-technical decision quality & trustworthiness", and Task 4.4 "AI Economics and regulation". The outcome of the risk identification and analysis process based on the proposed framework will result in adjusting and improving the functional and non-functional requirements. This is aligned with the EU AI Act and ISO demand for continuous monitoring and updates. In this way, we ensure that the improved understanding of the application will reflect in the product development, and the increasing level of technical maturity will be incorporated not only in documentation, but also for the practical implementation.

## 6.3 SAFETY AND ROBUSTNESS ASSESSMENT

Given the criticality of power grids, railway networks, and air traffic management infrastructures, ensuring that AI models function accurately under various conditions, including adversarial perturbations and environmental changes, is paramount. This is ensured by assessing AI-based systems' robustness, resilience, and reliability.

The methodology applied in Task 4.2 and described in deliverable D4.1 (Lemetayer et al., 2025), follows a structured approach based on standardisation frameworks such as ISO/IEC 24029-2 and the AI Act requirements. The first step involved risk identification and assessment, according to the methodology described in Section 6.2 where various technical risks are identified, including data drift and adversarial attacks. Vulnerabilities in different AI components, ranging from model inputs to decision outputs, were evaluated.

This risk analysis revealed that a common aspect of these three infrastructures is that the critical technical risk source is perturbations in the state or input space, such as noise or missing measurement data, or even deliberate disruptions introduced through cyberattacks. These perturbations are more frequent (e.g., information collected from different internal and external data sources) and have a higher potential impact, as they cannot be mitigated via model replacement or retraining. Another risk mentioned in the power grid – and that leads to one specific use case – but that could be applied to other infrastructures, is out-of-domain data, particularly when digital environments are used for the initial training of the AI-based decision system.

Once the risks are identified, adversarial datasets are generated to simulate cyberattacks, sensor failures, and environmental disruptions. This involves using perturbation agents that create controlled variations in AI inputs, allowing for a thorough examination of system performance under stress. Domain-specific perturbation models are being developed for power grids, railways, and air traffic control to ensure that AI models remain functional despite real-world uncertainties and are described in deliverable D4.1.

Then, multiple evaluation metrics are defined in deliverable D4.1 to measure robustness, resilience, and reliability under controlled perturbations and compare them with baseline operational scenarios.

**Robustness** evaluations provide critical insights into how AI models respond to stress conditions and include performance degradation under adversarial conditions, stability of decisions under input perturbations, and the rate of successful adversarial attacks.

**Resilience** metrics ensure that AI systems can recover from adverse events effectively and thus focus on the area between performance degradation and recovery curves, and maximum deviation from nominal performance.

**Reliability** evaluations focus on the long-term sustainability and operational consistency of AI models in real-world applications, by assessing mean time between failures, accuracy detecting out-of-distribution data, and adaptation time to environmental shifts.

Collectively, these evaluations contribute to developing AI solutions that are technically robust, resilient to external disruptions, and capable of maintaining performance integrity over extended periods. AI systems should be subjected to continuous adversarial testing to identify potential vulnerabilities before deployment. By incorporating adversarial dataset generation as a core component of the evaluation process, developers can pre-emptively address issues that may arise in real-world applications.

## 6.4 DECISION QUALITY AND ETHICS

As the AI4REALNET project is concerned with the use of AI methods and applications (such as digital assistants) by operators in high-stakes scenarios, it is important to evaluate not only the quality of AI-generated solutions, but also the properties of the complete socio-technical systems, including AI agents and humans in various roles. This emphasis on socio-technical properties is aligned with a number of dimensions defined in the EU framework for Trustworthy AI, such as Requirement #1 “Human Agency and Oversight”, Requirement #2 “Technical Robustness and Safety”, Requirement #4 “Transparency”, Requirement #5 “Diversity, Non-discrimination and Fairness”, and Requirement #6 “Societal and Environmental Well-being”), as well as multiple statements related to the uptake of human-centric and trustworthy AI technologies in EU AI Act (starting with Chapter 1, Article 1).

Deliverable D4.1 (Lemetayer et al., 2025) has defined and described the following evaluation objectives for the respective WP4, in particular Task 4.3 “Social-technical Decision Quality & Trustworthiness”, based on the project proposal, deliverable D1.1 (Bessa et al., 2024), and ongoing project work:

- Social-technical decision quality,
- AI acceptability, trust, and trustworthiness,
- Human-user experience,
- AI and human learning curves,
- Task allocation balance,
- Long-term consequences of AI-assistants.

Notably, these objectives stem from the requirement identification process performed during the definition of the use cases in WP1. This process adopted an approach of “trustworthiness-by-design" by using the structure of the ALTAI tool to identify requirements related to the trustworthy dimensions.

Most of the respective evaluation protocols and metrics will rely on objective measurements (such as participant task performance and psychophysiological measurements) and subjective data (such as questionnaires and interviews).

The first evaluation objective, “Social-technical decision quality”, puts the stress on the context of human operator interaction with the AI assistant, and considers the measurements of human intervention frequency in the automated decision-making process, the significance of such revisions, and the novelty of AI-generated decisions as assessed by the human operator. These concerns are

aligned with several requirements from ALTAI (such as Requirement #1 “Human Agency and Oversight” > “Human Oversight”, Requirement #2 “Technical Robustness and Safety” > “Accuracy”, and Requirement #4 “Transparency” > “Traceability”) as well in EU AI Act Chapter 3 (Section 2, Articles 13 et seq.).

The second evaluation objective, “AI acceptability, trust, and trustworthiness”, focuses on human operator-oriented indicators of i) trust and acceptability for the AI assistant in general, and ii) agreement with and trustworthiness of specific AI-generated decisions. Furthermore, the evaluation protocols defined for this objective also address the perceived decision comprehensibility/explainability, as the operators’ ability to understand the decision can affect their ability and willingness to act on it. All these concerns are aligned with several requirements from ALTAI (such as Requirement #4 “Transparency” > “Explainability” and “Traceability”) as well as in the EU AI Act (e.g., Chapter 3, Section 2, Article 15 or Chapter 3, Section 4, Article 50).

The third evaluation objective, “Human-user experience”, focuses on the respective concerns such as workload and stress, alignment with and support for the operators’ cognitive processes, and user motivation and satisfaction (including the operators’ satisfaction with the system’s support for their decision-making process). While this objective is overall oriented towards the human users rather than AI decisions, it aligns with several requirements from ALTAI (such as Requirement #1 “Human Agency and Oversight” > “Human Agency and Autonomy” and Requirement #6 “Societal and Environmental Well-being” > “Impact on Work and Skills”) as well in EU AI Act Chapter 3 (Section 2, Articles 13 et seq.).

The fourth and fifth evaluation objectives, “AI and human learning curves” and “Task allocation balance”, consider the co-learning and balance aspects in human-AI teaming. Therefore, they align with several requirements from ALTAI (Requirement #4 “Transparency” > “Communication” and Requirement #6 “Societal and Environmental Well-being” > “Impact on Work and Skills”) as well as the EU AI Act Chapter 3 (Section 2, Article 14 and 15).

Finally, the sixth evaluation objective, “Long-term consequences of AI-assistants”, focuses on perceived and predicted long-term consequences of AI assistant adoption, including the reflection on issues such as operator trust, agency, deskilling, overreliance, and the need for additional training for adopting the AI assistant. The objective also includes reflection on biased decisions potentially produced by the AI assistant with respect to gender/ethnicity/age, or commercial interests. Finally, the self-reported predicted adoption of the AI assistant by users, stakeholders, or experts is also included in this evaluation objective. These concerns are aligned with the statements in several requirements from ALTAI (such as Requirement #1 “Human Agency and Oversight” > “Human Agency and Autonomy” and “Human Oversight”, Requirement #4 “Transparency” > “Explainability”, Requirement #5 “Diversity, Non-discrimination and Fairness” > “Avoidance of Unfair Bias”, Requirement #6 “Societal and Environmental Well-being” > “Impact on Work and Skills”) as well as Chapter 3 in EU AI Act (Section 2, Articles 13 et seq.)

The activities starting within Task 4.3 and aligned with other Tasks within WP4 include continuous monitoring of the proposed solutions during the development and validation phases, including the assessment of use case requirements fulfilment, which informed the evaluation objectives described above with respect to decision quality and trustworthiness.

## 6.5 ETHICS CONSIDERATIONS IN AI4REALNET ALGORITHMS

WP2 ("Fundamental AI building blocks") develops AI algorithms and interfaces for safety-critical infrastructures, emphasizing transparency, robustness, and explainability. Specifically, Task 2.1 focuses on physics- and knowledge-informed AI, while Task 2.3 develops explainable, interpretable, and safe AI methods, including Human-Machine Interface (HMI) innovations. Together, they ensure alignment with the ALTAI ethical dimensions identified in WP1 and support trustworthy AI development in compliance with the upcoming EU AI Act.

Here, it follows the mapping of WP2 contributions to ALTAI ethical dimensions.

In terms of **human agency and oversight**, although WP2 centers on autonomous AI modes, it supports human oversight by developing knowledge-assisted models (T2.1) and explainable RL frameworks (T2.3). Task 2.3.2 further enhances automation transparency via cognitive-aware HMIs, enabling real-time understanding of AI goals, limitations, and confidence—crucial for effective human-in-the-loop and on-the-loop designs.

WP2 improves **technical robustness and safety** via multiple routes:

- Physics-informed AI (T2.1) increases model generalizability, especially under low-data or uncertain conditions.
- Safe RL (T2.3.1) ensures safer learning of control policies.
- Agent failure prediction and uncertainty quantification help detect unsafe conditions early. These elements contribute to trustworthy operation in dynamic, high-risk settings.

**Transparency** is a central objective in WP2, which develops interpretable learning models and explainability tools such as prototype learning, agent behavior analysis, and interactive visualizations (T2.3.1). HMI enhancements in T2.3.2 improve real-time interpretability, reinforcing transparency and facilitating auditability for operators and developers.

In terms of **diversity, non-discrimination, and fairness**, WP2 algorithms do not operate on human-centric data and thus are not expected to exhibit bias toward individuals. However, fairness risks may arise if these models are embedded in broader decision systems. WP2 flags this for consideration in WP4 and systems integration.

In terms of **societal and environmental well-being**, while WP2 does not directly address societal impacts, its contributions enable **efficient, reliable, and interpretable AI**, indirectly supporting societal goals (e.g., lower emissions through optimized control). HMIs designed with human-centered principles (T2.3.2) also promote safe operator interaction and well-being.

WP2 improves **accountability** by enabling **traceable decisions** via interpretable agents and visual explanations. System behavior can be logged and analyzed post hoc, and T2.3 methods support **auditability**, aligned with AI Act expectations for high-risk AI systems.

In WP3 ("AI-augmented human decision-making"), the emphasis is put on devising a co-learning architecture that fosters three control modes: 1) full human control and decision-making, supported by AI algorithms, 2) human-AI co-learning, and 3) full autonomous AI control directed by human operators. In all control modes, deliverable D1.1 (Bessa et al., 2024) identified that *meaningful* human

control is needed to influence AI-based recommendations and decisions, requiring human oversight and the (technical) capability to intervene or even bypass AI decisions.

In Task 3.1 of WP3, critical parameters underlying *human* decisions have been identified and listed, which serve as requirements for the implementation of AI algorithms supporting humans, human-AI data exchange (e.g., for co-learning), human intervention capabilities, and information needs for effective supervision. In Task 3.2, the emphasis is put on multi-objective optimization and decision making, underlining the need to carefully trade off and balance decisions against a multitude of system performance parameters, as well as criteria representing societal and human values. This not only requires that algorithms are sufficiently robust against uncertainties in such decision-making parameters, but that they are also flexible enough to consider human feedback (e.g., accept/reject/rate recommendations) and adapt their policies accordingly. In Task 3.3, co-learning aspects are considered that focus on how humans can learn from AI, and how AI can learn from humans. Co-learning strategies and requirements, identified as part of Task 3.1, are centered around humans acting as the final and responsible decision maker, while AI *supports* rather than replaces the human operator. In Task 3.4, full autonomous AI decision-making is considered, while the human operator remains in control of directing AI agents in task completions.

In summary, the activities in WP3 address the following concerns raised in WP1, and documented in deliverable D1.1, as follows:

**Human agency and oversight.** In Tasks 3.1-3.3, AI assists human operators in managing work and tasks by providing recommendations and advice (either automatically or by request), but human operators retain full control over decision-making and the execution of decisions. For this, human-machine interfaces (HMIs), developed and/or extended WP2 and WP3, will be used to: 1) enable human operators to exercise full control even without AI-based recommendations, 2) provide humans with insights into critical control parameters to judge the validity and impact of their own and AI-based solutions, and 3) nudge and/or revise AI-based recommendations in more acceptable directions in terms of societal constraints and values. In case of full autonomous AI control (T3.4), humans will use the HMIs to monitor AI behavior and can delegate tasks to AI agents (and take back control if needed).

**Technical robustness and safety**. In real-life operational settings, (data) uncertainty and unanticipated perturbations will most certainly be part of the control loop. The AI algorithms must therefore be sufficiently robust against uncertainties and system perturbations, which will be analyzed and modeled as part of Task 3.1. In addition, order-agnostic methods will be considered in Task 3.3 and can be considered part of increasing the flexibility of AI algorithms, as a counterpart to robustness.

**Transparency**. System transparency is crucial if effective human agency and oversight are required. This will be a part of Tasks 3.3-3.4, which employ the developed HMIs to provide explainability for understanding and monitoring AI solutions. This also includes bi-directional communication, allowing humans and AI to exchange data required for decision making and learning. For transparency purposes, these data structures will be recorded.

**Societal and environmental well-being**. AI systems are designed to find (near-optimal) solutions that have the potential to reduce carbon footprint (e.g., by reducing track miles in transportation systems) and increase resilience to extreme events (e.g., weather). This requires techniques that can balance multiple (and sometimes competing) objectives. In Task 3.2, such techniques (e.g., interactive Pareto Front visualizations) are developed and implemented in ways that enable human operators to 1)

inspect how different objectives have been balanced and 2) select solutions that have a different, yet feasible and acceptable, weighing.

## 6.6 ANALYSIS OF THE ETHICS AND DATA PROTECTION COMMITTEE

The Ethics and Data Protection Committee of the AI4REALNET project convened on 4 September 2025 to review the latest deliverable on ethical and data protection issues. The session was chaired by Ricardo Bessa and included data protection and ethics officers from project partners, with 12 representatives in the meeting: SBB, INESC TEC, ZHAW, Enlite.ai, TenneT, FLATLAND, TU Delft, University of Kassel, RTE, LiU, and IRT SystemX.

**Overall Assessment**

The committee acknowledged the substantial progress made in integrating ethical and data protection considerations into the project. Members highlighted the usefulness of the summary deliverable in consolidating inputs from various work packages.

**Data Protection and Privacy**

- Positive feedback was provided on the robustness of GDPR compliance measures.
- It was noted that the handling of datasets, acquisition processes, and informed consent procedures is thorough and well-documented.
- A recommendation was made to improve clarity on data flows within the consortium, particularly in cases involving biometric data or worker performance monitoring.
- Special attention was drawn to legal requirements at the national level, especially concerning health-related data that cannot be shared with employers.

**AI Ethics and Use Cases**

- The committee suggested clarifying the main objectives and decisions supported by AI systems in each use case to improve readability and avoid overlap with other deliverables.
- It was recommended to explain why certain ethical requirements (e.g., transparency) are prioritized in specific use cases.
- Concerns were raised about the potential for AI systems to monitor workers, with advice to ensure this aspect is carefully framed within ethical and legal boundaries. It was explained that in AI4REALNET there is not intention of having AI systems monitoring workers, in fact the whole concept is about having AI supporting workers in their decisions and learning also from their actions and feedback.

**KPIs and Risk Assessment**

- The project's approach of linking Key Performance Indicators (KPIs) with risks was welcomed as a tangible way of measuring ethical dimensions.
- While the high number of KPIs was acknowledged, it was agreed that this breadth is necessary given the diversity of infrastructures and domains involved.
- A proposal was made to include a summary table of objectives per requirement to complement the current framework.

The committee encouraged the consortium to continue strengthening reflections on human oversight, socio-technical impacts, and long-term implications of AI deployment in critical infrastructures.

**Conclusions and Next Steps**

- The committee concluded that this deliverable provides a solid and comprehensive foundation for addressing ethics and data protection in AI4REALNET.

- No major concerns were raised from the data protection perspective.
- It was reaffirmed that the project will maintain human oversight in all AI-supported decision-making processes, ensuring humans remain central in operational control.

# REFERENCES

Aguiar Castro, J. (2023). AI4REALNET: Data Management Plan structure and objectives. Zenodo. https://doi.org/10.5281/zenodo.10804943

Bessa, R.J., Yagoubi, M., Leyliabadi, M., et al. (2024, September). AI4REALNET framework and use cases. AI4REALNET Deliverable D1.1. [Online] https://ai4realnet.eu/wp-content/uploads/2024/12/D1.1-AI4REALNET-framework-and-use-cases_v1.0-3.pdf

Bessa, R.J. (2024). AI4REALNET Project presentation: Ethics and Data Protection Committee (EDPC) meeting. Zenodo. https://doi.org/10.5281/zenodo.10804373

European Commission (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Version 3. July, 2016

European Commission (2021). Identifying serious and complex ethics issues in EU-funded research. Version 5. July, 2021

ISO 31000. 2018. Risk Management Guidelines. Standard, International Organization for Standardization, Geneva, CH.

ISO/IEC 23894. 2023. Information technology - Artificial intelligence - Guidance on risk management. Standard, International Organization for Standardization, Geneva, CH.

Lemetayer, B., Saporetti, L., Schneider, M., Bessa, R.J., Liessner, R., et al. (2025, March). Evaluation and test protocols. AI4REALNET Deliverable D4.1. [Online] https://ai4realnet.eu/wp-content/uploads/2025/03/D4.1_Evaluation-and-test-protocols-v1.0.pdf

Stefani, T., Deligiannaki, F., Berro, C., Jameel, M., Hunger, R., Bruder, C., Krüger, T. (2023, October). Applying the Assessment List for Trustworthy Artificial Intelligence on the development of AI supported Air Traffic Controller Operations. In 2023 IEEE/AIAA 42nd Digital Avionics Systems Conference (DASC) (pp. 1-9). IEEE.

UNISDR. 2004. Living with risk: A global review of disaster reduction initiatives. United Nations International Strategy for Disaster Reduction.

# ANNEX 1 – DMP INFORMATION GATHER TEMPLATE

## DMP information – Partner

This document, based on the *The Data Curation Profiles Toolkit Interview Worksheet*[3] and the *Guidelines on FAIR Data Management in Horizon 2020*[4], aims to gather information to support the development of the DMP.

For any doubt, please contact: João Aguiar Castro (DMP manager), joao.a.castro@inesctec.pt

### Data Summary

**Please provide a description of the datasets you are expected to generate.** Please add as many lines and dataset information as you seem fit. Identify subtask if needed.

**Dataset description**

| Task | Dataset | General information |
|---|---|---|
| | Name of the dataset<br>(as detailed as possible)<br>The dataset name will be used throughout the form tables. | **Type of data:**<br>**How the data will be collected:**<br>**Software, instrument or tool required to process the data:**<br>**Expected size and format:**<br>**Update frequency:**<br>**Personal or Sensitive data:** |
| | | **Type of data:**<br>**How the data will be collected:**<br>**Software, instrument or tool required to process the data:**<br>**Expected size and format:**<br>**Update frequency:**<br>**Personal or Sensitive data:** |
| | | **Type of data:**<br>**How the data will be collected:** |

[3] Carlson, Jake (2010) The Data Curation Profile Toolkit: Interview Worksheet. Purdue University

[4] European Commission (2016) Guidelines on FAIR Data Management in Horizon 2020

| | | |
|---|---|---|
| | | **Software, instrument or tool required to process the data:**<br>**Expected size and format:**<br>**Update frequency:**<br>**Personal or Sensitive data:** |

**How the dataset relates to the project´s objectives?** A brief description to contextualize the dataset in the overall AI4REALNET objectives.

| Dataset | Aim and relation with the overall objective of the project |
|---|---|
| | |
| | |
| | |

## FAIR data

**Can you specify <u>when</u> and <u>under which conditions</u> the datasets can be shared within the project?** Taking the data lifecycle into account, it is expected that other project partners can access the data.

| Dataset | Conditions for sharing the dataset within the project |
|---|---|
| | |
| | |
| | |

**Can you specify <u>when</u> and <u>under which conditions</u> the datasets can be made publicly available?** For instance, whether the dataset can be made available in a data repository as soon as results are published (at the time of submission, or publication), or whether access conditions have to be defined (under embargo or request).

| Dataset | Expected availability of the dataset |
|---|---|
| | |
| | |
| | |

**Will the dataset be created with interoperability in mind? In which way?** If the datasets use formats and are collected with methodologies that promotes the integration with other data, applications or workflows. For instance, standardized disciplinary vocabularies, or other standardized practices, promotes interoperability.

| Dataset | Interoperability |
|---|---|
| | |
| | |
| | |

**What kind of documentation will be produced to enable easy access to and the use of the datasets?** Data documentation examples are readme.files, methodology information, codebooks, variables definitions, and others, any disciplinary standards to enable data sharing.

| Dataset | Needed data documentation |
|---|---|
| | |
| | |
| | |
| | |

**Are there any ethical or legal issues that can have an impact on data sharing? If so, what?**

| Dataset | Ethical or legal issue |
|---|---|
| | |
| | |
| | |
| | |

**Are you aware of any data policies of the journals where it is expected to publish results? Can you please name some of these journals?** This information can be used to identify possible editor's data policies to timely address them and include them in the DMP. For an example, see the Elsevier research data policy

## Data storage and organization

**Will the dataset be created or used in collaboration with different project partners?** If yes, fill in the table with the partners corresponding to each dataset.

| Dataset | Partners |
|---|---|
| | |
| | |
| | |

**What data storage and backup strategies will be adopted?** If there is any backup periodicity, how many copies of the datasets and what storage solutions (e.g. partners institutional drives combined with offline storage devices).

| Dataset | Storage and backup strategy |
|---|---|
| | |
| | |
| | |

**What are the most important parts of the data to preserve (manage and maintain over time) and under which conditions?** Not all the data needs to be preserved or published in a data repository. Does a particular dataset have future reuse value? Is the processed data more critical than the raw data?

| Dataset | Need for preservation and conditions |
|---|---|
| | |
| | |

**Who are the people responsible for data management?** For instance, the people responsible for systematically ensure that the data is fit for use. A point of contact within the WP (or task) to data management related issues.

| Dataset | Responsible for the dataset management |
|---|---|
| | |
| | |

## Other

**Any additional notes you may think necessary to provide more information about your datasets.**

# ANNEX 2 – DIGITAL ENVIRONMENTS DESCRIPTION

### Grid2Op: power grid

RTE developed the open-source Grid2Op environment to model and study a large class of power system-related problems and facilitate the development and evaluation of controllers (or agents) that

act on power grids. Any type of control algorithm in interaction with a virtual version of the electrical grid can be used to overcome gaps between research communities.

Through different L2RPN competitions, calibrated virtual environments have been instantiated for testing over robustness to adversarial attacks, adaptability for increasing renewable energy share, or agent alert trustworthiness. Such "autonomous" agent scenarios can already be visualized and analyzed through the Grid2viz module. Moreover, it is also possible for a human to play live scenarios, assisted by an AI agent with the Grid2Game module. A human can choose contextual triggers for alerts and get recommendations from the agent. It can also do its manual simulation if needed. Ultimately, this will be run through the OperatorFabric Hypervision Interface in real operations. There is also a repository hosting a set of reference baselines.

Chronix2Grid package allows the generation of synthetic but realistic consumption, renewable production, electricity loss (dissipation) and economic dispatched productions chronic for a given power grid.

**Flatland: Railway**

The Flatland environment is a comprehensive framework developed (by industry partners like SBB, DB, and AI community) for easy development and experimentation on the vehicle rescheduling problem for railway networks. Flatland represents railway networks as 2D grid environments with restricted transitions between neighboring cells. On the 2D grid, multiple train runs must be performed for a given set of goals and circumstances. Trains are represented as agents that make decisions on movement and navigation.

Flatland is a discrete-time simulation, i.e., it performs all actions with constant time steps. A single simulation step synchronously moves the time forward by a constant increment, thus enacting exactly one action per agent. The Flatland environment is tailored towards RL. It provides observations and rewards to any controlling agent, and it expects one discrete action per agent per step. Flatland, in its current state, provides a set of global and local observations. It provides generators for generating railway networks and demand for trains (scenarios), an evaluation system, and mechanisms to inject disturbances into rail operations. These disturbances are represented as malfunctions of trains, i.e., trains being unable to move on the track for several time steps. The occurrence of these is distributed according to configurable distribution at scenario definition.

After three competitions with Flatland, a comprehensive set of basic (mostly) AI solutions for Flatland exists and can be used as baseline/benchmark models.

**BlueSky: Air traffic management**

The BlueSky environment is an open-source ATM simulator that has been developed since 2013 with TU Delft as its main developer. It contains open source, open data aircraft performance models and a global navigation database including airports; it is also compatible with Base of Aircraft Data (BADA) v3.xx files (containing performance and operating procedure coefficients for 295 different aircraft types). Although BlueSky started out as a simulator aimed at conventional aviation, in recent years it has been extended with several Drone/Urban Air Mobility models and functionality and has since been applied in several UAM/UTM-related projects.

Through its modular setup, an extension of each of the components of BlueSky (e.g., autopilot, FMS, performance model, conflict detection and resolution, environmental modeling, visualization, etc.) can be reimplemented or extended. In the same way, it is also possible to add completely new functionality

to the simulator. By default, BlueSky has its own Qt/OpenGL-based interface that allows the user to control the simulation and get an overview of the simulated traffic. Through its client/server network implementation, BlueSky can also easily interface with separate ATM user interface applications and piloted blip driver stations.

# ANNEX 3 – INESC TEC DRIVE DESCRIPTION

Technical and security measures met by INESC TEC drive ensure:

- Deny unauthorized persons access to data processing equipment used for processing personal data (equipment access control). Nextcloud includes edition/collaboration tools that enable users to process data on the server side. Only authorized users can access, create, store, or edit those files. Computers at INESC TEC may be used for local data processing; these computers are restricted to users with INESC TEC accounts; each user has a local area without administration privileges to prevent them from accessing areas from other users.
- Prevent the unauthorized reading, copying, modification or removal of data. The Nextcloud instance enforces access control at folder or file levels to authorized users only. The permission control may restrict users from deleting or modifying data.
- Prevent the unauthorized input of data and the unauthorized inspection, modification or deletion of stored data (storage control). Data storage cannot be accessed directly by users, only through the Nextcloud instance, which prevents unauthorized input of data and the unauthorized inspection, modification or deletion of stored data. Servers hosting Nextcloud are hosted on a datacenter at INESC TEC with biometric access control, restricted to IT staff.
- Prevent the use of automated data processing systems by unauthorized persons using data communication equipment (user control). The Nextcloud instance prevents automated data processing by unauthorized persons.
- Ensure that it is possible to verify and establish to which bodies data have been or may be transmitted or made available using data communication equipment (communication control). The INESC TEC Nextcloud instance ensures traceability via activity logs, such as downloads, modification, access, data sharing, including automatic e-mail notifications.
- Ensure that it is subsequently possible to verify and establish which data have been input into automated data processing systems and when and by whom the data were input (input control).
- Ensure that the functions of the system perform without fault, that the appearance of faults in the functions is immediately reported (reliability) and that stored data cannot be corrupted by means of a malfunctioning of the system (integrity). The Nextcloud instance is hosted on servers being actively monitored, with redundant power supplies connected to UPS, and with redundant network connectivity. The Nextcloud instance provides file versioning and recover from data file deletion; moreover, the data is stored on a filesystem with file checksum, on a storage server with redundant disk parity, redundant power supplies connected to UPS, with redundant network connectivity, and with daily backups.

Data on the Nextcloud instance can only be inserted, stored, accessed, inspected, modified, or deleted by authorized users only.

Data transfers between the Nextcloud instance and user client computers/applications is done over secure, encrypted/authenticated communications (HTTPS).

Servers hosting Nextcloud are hosted on a datacenter at INESC TEC with biometric access control restricted to IT staff. The Nextcloud instance provides file versioning and recovery from data file deletion; the data is stored on a filesystem with a file checksum and daily backups. Therefore, depending on the level of data recovery needed, data can be recovered from the mechanisms provided by Nextcloud or from external backups. Besides short-term data backups, data can be stored on a long-term encrypted tape archive, with a second copy off-site for disaster recovery.

# ANNEX 4 – INFORMED CONSENT

**INFORMED CONSENT TO PARTICIPATE IN A RESEARCH PROJECT**

Please read the following information. If you think something is incorrect or unclear, don't hesitate to ask for more information through the email: (Duarte Filipe Dias duarte.f.dias@inesctec.pt)

If you wish to participate in the study, we request your consent, which you can give by signing the document.

Participation in the study is voluntary. You can quit it at any time without any consequence, needing only to contact the person responsible through the email specified above.

***1.PROJECT DESCRIPTION***

**Title:** AI4REALNET human-in-the-loop integration

**Responsible Entity:** INESC TEC

**Principal Investigator:** ***Duarte Dias*** - **duarte.f.dias@inesctec.pt**

**General Description:** The Center for Biomedical Engineering Research (C-BER) at INESC TEC has been developing for 10 years several wearable technologies focused on the monitoring of first responders in hazardous environments and operators in stressful positions, an area that is integrated in the quantified occupational health research line . Besides the focus on wearable devices, this research line has also made a strong effort to combine algorithms based on physiological signals with psychological information, to understand the impact of psychosocial risks, such as stress in user's health, using psychophysiological information. Our wearable devices are capable of monitoring physiological signals, namely electrocardiogram (ECG), respiration and temperature and jointly with our algorithms and psychological information we can extract information from stress and cognitive levels in near-real-time.

The current study is integrated in the **AI4REALNET European project** (Horizon Europe; 101119527) with the aim to research innovative methodologies for the integration of human psychophysiological conditions in new AI systems that help in the decision making of the operator of critical infrastructures. The project is developing these new AI agents that aim to support the operator, and this study aims to integrate the human psychophysiological information into these new AI agents, providing then information from the operator in a seamless and implicit way – e.g. the operator is starting to be stressed and the system, without the operator state it of inform the AI agent, the system is able to reduce the number of tasks, or tries to solve more issues to try to reduce the amount of stress from the worker. This implicit knowledge must be personalized for each operator and that for that a first experimental test needs to be conducted. After personalization, the system will be able to retrieve stress and cognitive levels while the operator is being monitored and during the use of digital environments from three different domains: electric power grid, railway and air traffic management. The information of stress is binary (stress and no stress) with a level of confidence based on the models accuracy and the information of cognition is normalized between 0 and 1.

***2.PROCESSING OF PERSONAL DATA***

**Objectives:** The current study aims to understand the user psychophysiological changes (e.g., stress and cognitive levels) during the different trials that will be made.

- **Procedure:** Participant will be requested to collect data during 3 periods according to the methodology defined: 1) baseline data collection procedure – data collected in a peaceful location; 2) standardized laboratory stress procedure – Trier Social Stress Test - TSST, along with a 2-Choice Reaction Time Task; 3) Operator monitoring in real-time during the use of digital environments. The following systems, fully controlled by INESC TEC, will be used during these procedures:

  1) **VitalSticker** that measures ECG, Heart Rate, Respiration, core temperature estimation, activity and posture based on a chest band with textile conductive electrodes (depending of the scenarios it might be needed gel electrodes).

  2) **Smartphone**: will contain an App to easily connect to the devices and place it in the user's pocket. The application does not aim to be used during the test; it just needs to be initiated to collect all the information from the devices and then stopped and the end of the tests.

  3) **Surveys:** a sociodemographic questionnaire will be used in the 1st and 2nd periods, along with a stress and fatigue scale. The sociodemographic questionnaire should be filled at the beginning of the tests and the scale at the beginning and end of tests. Other questionnaires are also going to be used during the 3rd period, namely: assistant disturbance (scale 0-5), workload based on NASA-TLX methodology and Impact on workload based on 7-point Likert scale.

  4) **Computer:** only used on standardized laboratory stress procedure by the research team to perform the some tasks.

**Personal Data:**

- ECG, Heart Rate, Respiration, Core temperature estimation, activity and posture will be collected with Vitalsticker;
- Surveys to collect sociodemographic, stress, and fatigue levels information, such as other information regarding the interaction with the digital environments such as workload and disturbance.

**Purposes of Personal Data Processing:** The data collected will be processed in accordance with the applicable national and EU legislation and will only be used by the researchers for scientific research purposes in the domains of quantified occupational health and AI decision support systems.

**Data Controller(s):** INESC TEC – Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Campus da FEUP, Rua Dr. Roberto Frias, 4200 - 465 Porto.

**Confidentiality, Security Measures and Data Retention Period:** All data acquired will be analysed for scientific purposes within the framework of the study in question; all those involved in this study have a commitment to confidentiality and non-disclosure of data or personal information taken from it; the data collected will be pseudonymised and aggregated at a later stage so that your individual participation in the study will not be identified. Some steps will be taken to protect your privacy, namely: a) each participant will be assigned a randomized code; b) all data collected will be stored on safe local servers at INESC TEC with restricted, controlled access, limited only to members of the research group involved; c) the data output (stress and cognitive levels) will be shared within the project partners using a secure protocol; d) the data

collected is intended to be anonymised after 6 months, at which point all information that could directly or indirectly identify participants will be destroyed.

**Personal Data Sharing:** Only anonymous and aggregated results may be disseminated/published in scientific publications, which may involve research teams from different institutions.

**Incidental Findings:**

In addition, if any problems or physiological anomalies are detected in the data collected, would you like to be informed?

Please indicate: _____YES _____ NO

If you answer yes, please indicate your contact details (email or phone):_____.

**Rights of the Data Subject:** As the subject of the data, the law grants you the following rights: Information, Access, Rectification, Portability, Erasure, and Restriction of processing. In the event of withdrawal, there is no prejudice to the processing of data collected up to that point, but you may withdraw your consent and thus request the deletion of your personal data, provided that you exercise this right before the respective and irreversible anonymisation takes place, which will occur after 6 months. To exercise any of your rights, please use the following e-mail address: duarte.f.dias@inesctec.pt.

**Data Protection Officer:** For any questions regarding the processing of your personal data, please contact our Data Protection Officer at: dpo@inesctec.pt.

Under the terms of the Article 77 of the GDPR, the data subject also has the right to lodge a complaint with a supervisory authority in the European Union. In Portugal, the supervisory authority is the CNPD (www.cnpd.pt).

***3. INFORMED CONSENT FORM***

| | |
|---|---|
| 1. I read and understood the information about the project, including the identity of the controller, the type of data collected, the purpose of the collection and the respective processing. | ☐ |
| 2. I read and understood the information about how the data is stored and for how long, including what will happen to my data in the case of ceasing my participation in the project. | ☐ |
| 3. I have been given the opportunity to ask questions and to clarify any doubts about the project. | ☐ |
| 4. I understand I can quit participating in the project at any point, without needing to provide any justification and without suffering any penalty or having my motives questioned. | ☐ |
| 5. I understood how to communicate my decision to quit, as well as how to exercise my rights as the data subject. | ☐ |

***The Participant:***

*I declare I have read and understood this document, as well as the verbal information that have been given to me previously. As such, I accept to participate in this study and allow the use of the data I offer voluntarily.*

Name: … … … … … … … ..... … … … ..... … … … … … … … … … … … …

| Signature: … … … … … … … ……. … … … … … … … … … … … … … … … … … … … … … … | Date: …… /…… /……. |
|---|---|

# ANNEX 5 – USE CASES SUMMARY DESCRIPTION

## POWER GRID

***Business problem:*** Electricity networks are transforming as the ongoing decarbonization and digitalization introduce clean generation technologies, electrify demand, enable demand-side flexibility, and digitize and/or add new devices. This directly impacts supervision systems in control rooms, which have to a point where they are no longer cognitively manageable. Networks are also aging, and infrastructure developments are more limited, yet they integrate more automata. AI can help to address more numerous, complex, and coordinated decisions, increasing uncertainty, overcrowded and fragmented work environments with multi-screen applications, and increasing human operator cognitive load.

***Key stakeholders:*** Transmission system operators (TSOs), human operators, transmission grid users, and electricity market participants.

**UC1.Power Grid: AI assistant supporting human operators' decision-making in managing power grid congestion**

***Objectives:*** The goal of a TSO, and thus human operators in the control room, is to control electricity transmission on the electrical infrastructure (transmission grid) while pursuing multiple objectives, firstly to keep the system state within acceptable limits and:

- Safely manage overloads on the electrical lines and, more specifically, remedial action recommendations;
- Make the most of the renewable energies installed by limiting the emergency redispatching call to thermal power plants emitting greenhouse gases;
- Ease the workload of the human operator needed to fulfill his/her missions;
- Integrate explainability, transparency, and trust considerations for the human operator.

***UC short description:*** The AI assistant oversees the transmission grid, using SCADA data and Energy Management System tools to identify issues and categorize them for human intervention. It monitors power flows, adhering to defined operational conditions. Anticipating problems, it sends alerts to the operator with confidence levels, avoiding excessive alerts to maintain operator focus. Action recommendations include topological changes, re-dispatching, and renewable energy curtailment. The human operator selects an action or seeks more information, exploring alternatives. After the operator decides, the AI assistant provides feedback through load flow calculations and logs decisions for continuous learning and interaction improvement. This UC only addresses congestion issues, even if

other types of issues can arise on the Transmission Grid and are handled by the operators (e.g., voltage values outside prescribed upper/lower limits).

***System description and role of the human operator:*** This UC describes an AI assistant that provides a human operator with recommendations for actions and/or strategies, considering the abovementioned objectives. The AI assistant shall also act in a "bidirectional" manner, i.e., capitalize on the actions and the feedback from the operator with a continuous "online" learning process. Different modes of interaction between AI assistants and human operators are possible, ranging from "full human control" to "full AI control." The selected mode depends on the industry domain and context. In this UC, an ex-ante choice is made to apply a hybrid interaction where the human operator gets the final word on AI assistant recommendations.

***Key benefits and impact of AI:*** Minimize operational costs; facilitate energy transition by reducing renewable energy curtailment and improving carbon intensity of actions; reduce the workload of the human operator; increase resilience to extreme (natural and man-made) events.

**UC2.Power Grid: Sim2Real, transfer AI-assistant from simulation to real-world operation**

***Objectives:*** Assess the capability of an AI assistant to be used for the operation of a "real" transmission grid, in the sense that the "real" environment does not exactly behave as the one available to the agent (that is implemented in the AI assistant) during training and simulation procedures, even if they share the same functional properties (same grid components and topology), and operational constraints. The main objectives are:

- Look at additional technical considerations to successfully deploy an AI assistant in the real world besides its sole ability to find solutions to simulated situations.
- Improving human trust when such systems are deployed in real-world environments.
- Allowing for iterative human-AI refinements with human feedback and insights.

***UC short description:*** Outlines two paths for an AI assistant to manage a transmission grid. 1) In coping with real-world conditions, the AI assistant monitors grid situations, raises alerts for human intervention, and provides action recommendations, considering uncertainty from noisy and partially missing data. The human operator makes decisions based on AI suggestions, with feedback loops to continuously improve interactions and learn from realized actions. 2) When data limitations prevent full autonomy, the AI assistant alerts the human operator due to missing or poor-quality data. The operator can provide missing information to aid the AI in such cases. Enriched context, including human input and decisions, is logged for continuous learning, enhancing the AI assistant's robustness in making recommendations for grid actions.

***System description and role of the human operator:*** The AI assistant can still recommend actions to the human operator even with lower-quality data than used in training. However, this data may not enable fully autonomous recommendations, requiring the AI to seek additional feedback from the operator and raise an inaccuracy alert. When the AI cannot evaluate the need for action or a recommended action fails to produce the expected outcome, the operator can provide specific missing information to assist the AI in forecasting system states and assessing recommendations. As for the AI-assistant training, the human operator's decision and perception will rely on "theoretical simulations" (training and simulation tools).

***Key benefits and impact of AI:*** Minimize operational costs; facilitate energy transition by reducing renewable energy curtailment and improving carbon intensity of actions; reduce the workload of the human operator; increase resilience to extreme (natural and man-made) events.

### RAILWAY NETWORK

***Business problem:*** Growing environmental awareness and changing policies for mobility will lead to considerably more demand from railway network capacity, denser traffic, and a further need for efficiency and resilience of railway traffic management. Novel dispatching technologies or huge infrastructure investments are inevitable to maintain or improve the current quality of services. AI-based support systems can be developed to enhance dispatchers' capabilities, aiming to automate some of today's decision-making processes and provide support and input for human decision-making in complex operating scenarios.

***Key stakeholders:*** Railway network operators, network supervisors, railway undertaking operation managers, passengers, government, and society.

**UC1.Railway: Automated re-scheduling in railway operations**

***Objectives:*** The system's objective is to fully automate re-scheduling in railway operations to fulfill all offered services and minimize delays for the customer (passenger).

***UC short description:*** Unexpected events, such as infrastructure malfunctions or delays, can occur in railway operations. In this case, the automated system must re-calculate the schedule so the requested services can be fulfilled with as little delay as possible. Adapting the schedule includes interventions, such as changing the speed curves of trains, changing the order of trains at the infrastructure element, changing the routes of trains, or changing the platform of a commercial stop at a station. An automated AI-based system is designed to manage and optimize railway schedules in real-time, ensuring efficient rail network use while minimizing passenger delays. The system is constantly monitored by a human operator who can adjust the system's configuration and identify the need for adaptation and re-training.

***System description and role of the human operator:*** An AI-based re-scheduling system performs the re-scheduling task in a highly automated manner. This system observes the real-time state of all the trains and tracks in the control area of interest and automatically detects the need to intervene, decides on an intervention, and executes this intervention. Such an AI system for highly automated re-scheduling in operations is something new and unusual. The approach followed here is a first step towards introducing such a system. The highly automated AI system is treated as a new tool that is supervised and evaluated by an expert. In operations, the AI system re-schedules in a fully automated manner while the human supervisor monitors:

- The system's state in operations (e.g., number of trains, potential bottleneck in current and planned network usage)
- KPIs for the actual situations (e.g., current delay)
- Confidence/certainty of the AI system
- Intensity of intervention (how much changes to the current operational plan did the AI perform, e.g., change platform)
- The supervisor uses this information to:

- Decide at which point it would be advisable to switch off the AI system and take over control.
- Decide to re-configure/adjust the system in operations.

***Key benefits and impact:*** Improve punctuality of trains; increase the speed in response to disruptions or changes; better use of the available capacity in the railway network.

**UC2.Railway: AI-assisted human re-scheduling in railway operations**

***Objectives:*** Aims to use AI-based methods to assist the human dispatcher in railway operations in re-scheduling train runs to fulfill all offered services and minimize delays for the customer (passenger).

***UC short description:*** An AI-assistant system supports the human dispatcher. This system receives the real-time state of all the trains and tracks in the dispatcher's control area and derives possible dispatching options in case of deviations from the pre-planned schedule due to disruptions or delays. The options are presented in near real-time to the dispatcher and consist of actions the dispatcher can perform to bring the trains back or close to their pre-planned schedules. At any time during operations, the human-AI team can detect an emerging deviation of the actual state of the system from the planned state. The re-scheduling process can be initiated by various triggers such as infrastructure changes, train delays, equipment malfunctions, or potential future issues. The system is designed to detect these deviations in real-time and assess their impact on the overall schedule. The system also predicts issues that might become relevant in the future. The human learning process (e.g., to detect emerging deviations or to develop solutions) is explicitly supported by human-AI interaction.

***System description and role of the human operator:*** The human provides feedback (e.g., context unknown to the system), which is used by the AI to adapt the solutions. The human agent can choose to select one of the suggestions provided by the AI systems, initiate a new solution search, or choose their own course of action. Alternatively, humans formulate a hypothesis, and the AI system provides evidence for and against these hypotheses. Moreover, a human supervisor reviews the system's performance, analyzing how effectively it responded to deviations and the impact on service delivery. Based on this review, adjustments are made to the system's parameters, such as altering the prioritization criteria, adjusting acceptable delay thresholds, or refining the algorithm for schedule recalculations.

***Key benefits and impact:*** Improve punctuality of trains; increase the speed of response to disruptions or changes; better use of the available capacity in the railway network.

### AIR TRAFFIC MANAGEMENT

***Business problem:*** Air traffic density in European airspaces is steadily increasing. At the same time, pressing economic and environmental concerns force a fundamental shift towards time- and trajectory-based air traffic operations. Taken together, increased traffic loads and operational complexities may eventually drive the workload peaks of the tactical air traffic controller (ATCO) beyond acceptable thresholds, threatening the overall safety of the ATM system and hindering a smooth transition toward a sustainable future of ATM. Furthermore, for instance, in the Lisbon Flight Information Region (FIR), serviced by NAV Portugal, operational complexities arise from the activation of military areas, which can significantly restrict the usage of the upper airspace for General Air Traffic, requiring traffic to deviate horizontally, especially when in combination with unexpected events.

***Key stakeholders:*** ATC and Flow Management Position (FMP) staff manager/supervisor, air navigation service provider (ANSP) responsible for the flight information region, tactical air traffic controller, airlines, and pilots.

**UC1.ATM: Airspace sectorization assistant**

***Objectives:*** To partially and fully automate the sectorization process to assist or replace the ATC supervisor in deciding when and how to split and merge sectors to balance the workload of tactical ATCOs.

***UC short description:*** At ATC Centers, an operational supervisor exclusively decides when and how to split and merge sectors best, warranted by situational demands and available ATCO personnel. The degrees of freedom in sectorization involve considering horizontal (2D geometry) and/or vertical (altitude) constraints and can thus result in sectors split horizontally and/or vertically. Under nominal conditions, the supervisor typically can install several pre-fab sectorization options. However, unexpected events, such as deteriorated weather conditions, flight emergencies (e.g., aircraft equipment failure), and unscheduled ATC personnel shortages (e.g., due to sickness), may require non-standard sectorizations to be installed. An AI assistant, capable of operating under various levels of automation, will provide recommendations or even execute decisions on splitting the sector best horizontally, vertically, or both to balance the ATCO workload while ensuring safety and efficient traffic flows. It will also act bidirectionally by allowing the human operator to nudge the AI-generated recommendations in more favorable directions.

***System description and role of the human operator:*** The system automatically observes the real-time data from all relevant ATM platforms, predicts how and when to sectorize, and implements prediction results either as recommendations (to the human supervisor) or automatically installs the sectorization plan. The AI system can be considered a new tool supervised and evaluated by a human expert. The AI system communicates its decisions on an auxiliary display that, for example, visualizes sector configurations on a map-like interface. At lower levels of automation, the role of the human operator (here, the ATC supervisor) is to evaluate the AI-based recommendations by requesting additional information and explanations, accepting or rejecting advisories, and nudging AI decisions in a different direction by manual interventions. All decisions and interactions will be logged, allowing the AI system to learn from human preferences continuously. At higher levels of automation, the AI recommendations are executed based on “management by consent” (= AI implements only when the human accepts) or “management by exception” (= AI implements, unless the human vetoes). At the highest level of automation, the AI system is automatically implemented, and humans can only revise the system’s decisions afterward.

***Key benefits and impact of AI:*** Facilitate continuing growth of air traffic demand while maintaining high safety. Improve predictability of a certain sectorization over a certain time horizon.

**UC2.ATM: Flow & airspace management assistant**

***Objectives:*** The system's objective is related to the flight execution phase when a military area is activated, and the ATC must issue deviations to avoid the activated area. The goal is to recommend deviations with better sector capacity adherence and performance measured by an indicator of the environmental area – *en-route flight inefficiency of the actual trajectory*. The UC also considers the

need to review the sectorization plan due to the activation of military areas and the required trajectory-efficient deviations.

***UC short description:*** Some airports' activation/deactivation of military airspace can induce deviations from the flight plan routes. In this sense, to optimize the lateral deviation of the flights due to avoidance of an eventual temporary military-activated area, an AI assistant can analyze and suggest a decision in sectorization and routing of the main flows in the FIR. Human operators, more specifically the ATC and FMP supervisors, will be supported by an AI assistant in determining how to configure airspace sectors best and optimize the routes for traffic flows in the en-route sectors of the FIR. The AI assistant will also act bidirectionally by allowing the human operator to nudge the AI-generated recommendations in more favorable/acceptable directions. The airspace sectorization and flow structures, as devised by the AI and nudged by the operators in the pre-tactical phase, will be used by the tactical ATCO to manage traffic around the military-activated areas.

***System description and role of the human operator:*** An AI-based system highly automates the airspace design for capacity and flow management for operational scenarios. This system automatically observes data from all relevant ATM platforms, predicts how to organize the airspace regarding routings and sectorization, and implements results as recommendations to the human operator (e.g., ATC and FMP supervisors). The AI system can be considered a new tool that is supervised and evaluated by a human expert. The AI system communicates its decisions on an auxiliary display that, for example, visualizes airspace configurations on a map-like interface. The role of the human operator (here, the ATC and FMP supervisors) is to evaluate the AI-based recommendations by requesting additional information or explanations, accepting or rejecting advisories, and nudging AI decisions in a different direction through manual interventions. All decisions and interactions will be logged, allowing the AI system to continuously learn from human preferences.

***Key benefits and impact of AI:*** Facilitate continuing growth of air traffic demand while maintaining high safety. Improve a key performance environment indicator based on actual trajectory, measuring the average en-route additional distance concerning the great circle distance.