**Lab Project Evaluation Sheet:**

Classify your AI-based application concerning the system autonomy.

Justify your statements! They need to be Understandable and comprehensible argumentations.

Classify your AI-based application concerning the safety criticality. Justify your statements.

Perform the following steps based on the EASA Concept Paper: First usable guidance for Level 1 & 2 machine learning applications

1. Define and Specify Application
2. Perform a trustworthiness analysis
3. Consider the AI Assurance for you Application
4. Consider the Human Factors for AI
5. Perform an AI safety risk mitigation
6. Consider organizational changes related to the AI application

Describe how you would approach the different objectives for you AI application and justify your answers.

**Example:**

Objective: HF-09

The applicant should design the AI-based system with the ability to negotiate, argue, and support its positions.

Answer: *Not applicable as there is no user-interaction within the AI system.*

*In order to select the right trajectory, the flight crew will have several options to choose from*

1. Trustworthiness analysis
   1. Characterization of the AI Application
      1. High-Level task(s) and AI-based system definition
      2. Concept of operations for the AI application
      3. Functional Analysis of the AI-based system
      4. Classification of the AI application
   2. Safety Assessment of ML Applications
      1. AI safety assessment concept
      2. Impact assessment of AI introduction
      3. Initial safety (support) assessment
      4. Instruction for continuous safety assessment
   3. Information security considerations for ML applications
   4. Ethics-based assessment
2. AI assurance
   1. Learning Assurance
      1. Learning Assurance process planning
      2. Requirements and architecture management
      3. Data Management
      4. Learning process management
      5. Training and learning process validation
      6. Learning process verification
      7. Model Implementation
      8. Interference Model verification and integration
      9. Data and learning verification
      10. Verification of (sub)system requirements allocated to the AI/ML constituent
      11. Configuration Management
      12. Quality and process assurance
   2. Development & post-ops AI explainability
3. Human Factors for AI
   1. AI operational explainability
      1. Target Audience for operational explainability
      2. Need for operational explainability
      3. Anticipated operational explainability modulation
      4. Objectives for operations AI explainability
   2. Human-AI-Teaming
   3. Modality of interaction and style of interface
   4. Error Management
      1. Contribution of AI-based systems to a new typology of human errors
      2. How AI-based system will affect the methods of error management
   5. Failure Management0
   6. Additional Topics under development
4. AI safety risk mitigation
   1. AI safety risk mitigation concept
   2. AI SRM top-level objectives
5. Organizations
   1. High-level provisions and anticipated AMC
   2. Competence considerations
   3. Design Organization Case
6. **Trustworthiness analysis**
   1. **Characterization of the AI Application**
      1. **High-Level task(s) and AI-based system definition**

**Objective CO-01:** The applicant should identify the list of end users that are intended to interact with the AI-based system, together with their roles, their responsibilities and their expected expertise (including assumptions made on the level of training, qualification and skills).

**Objective CO-02:** For each end user, the applicant should identify which high-level tasks are intended to be performed in interaction with the AI-based system.

**Objective CO-03:** The applicant should determine the AI-based system taking into account domain-specific definitions of ‘system’.

* + 1. **Concept of operations for the AI application**

**Objective CO-04:** The applicant should define and document the ConOps for the AI-based system, including the task allocation pattern between the end user(s) and the AI-based system. A focus should be put on the definition of the OD and on the capture of specific operational limitations and assumptions.

**Objective CO-05:** The applicant should document how end users’ inputs are collected and accounted for in the development of the AI-based system.

* + 1. **Functional Analysis of the AI-based system**

**Objective CO-06:** The applicant should perform a functional analysis of the system.

* + 1. **Classification of the AI application**

**Objective CL-01:** The applicant should classify the AI-based system, based on the levels presented in the Table, with adequate justifications.

* 1. **Safety Assessment of ML Applications**
     1. **AI safety assessment concept**
     2. **Impact assessment of AI introduction**
     3. **Initial safety (support) assessment**

**Objective SA-01:** The applicant should perform a safety (support) assessment for all AI-based (sub)systems, identifying and addressing specificities introduced by AI/ML usage.

* + 1. **Instruction for continuous safety assessment**

**Objective ICSA-01:** The applicant should identify which data needs to be recorded for the purpose of supporting the continuous safety assessment.

* 1. **Information security considerations for ML applications**

**Objective IS-01:** For each AI-based (sub)system and its data sets, the applicant should identify those information security risks with an impact on safety, identifying and addressing specific threats introduced by AI/ML usage.

**Objective IS-02:** The applicant should document a mitigation approach to address the identified AI/ML-specific information security risk.

**Objective IS-03:** The applicant should validate and verify the effectiveness of the security controls introduced to mitigate the identified AI/ML-specific information security risks to an acceptable level.

* 1. **Ethics-based assessment**

**Objective ET-02:** The applicant should ensure that the AI-based system bears no risk of creating human attachment, stimulating addictive behaviour, or manipulating the end user’s behaviour.

**Objective ET-03:** The applicant should ensure that the AI-based system presents no capability of adaptive learning.

**Objective ET-04:** The applicant should comply with national and EU data protection regulations (e.g. GDPR), i.e. involve their Data Protection Officer (DPO), consult with their National Data Protection Authority, etc.

**Objective ET-05:** The applicant should ensure that procedures are in place to avoid creating or reinforcing unfair bias in the AI-based system, regarding both the data sets and the trained models.

**Objective ET-06:** The applicant should perform an environmental impact analysis, identifying and assessing potential negative impacts of the AI-based system on the environment and human health throughout its life cycle (development, deployment, use, end of life).

**Objective ET-07:** The applicant should define measures to reduce or mitigate the impacts identified under **Objective ET-06.**

**Objective ET-08:** The applicant should identify the need for new skills for users and end users to interact with and operate the AI-based system, and mitigate possible training gaps (link to **Provision ORG-06, Provision ORG-07**).

**Objective ET-09:** The applicant should perform an assessment of the risk of de-skilling of the users and end users and mitigate the identified risk through a training needs analysis and a consequent training activity (link to **Provision ORG-06, Provision ORG-07**).

1. **AI assurance**
   1. **Learning Assurance**
      1. **Learning Assurance process planning**

**Objective DA-01:** The applicant should describe the proposed learning assurance process, taking into account each of the steps described in Sections C.3.1.2 to C.3.1.12, as well as the interface and compatibility with development assurance processes.

* + 1. **Requirements and architecture management**

**Objective DA-02:** Documents should be prepared to encompass the capture of the following minimum requirements:

* safety requirements allocated to the AI/ML constituent;
* information security requirements allocated to the AI/ML constituent;
* functional requirements allocated to the AI/ML constituent;
* operational requirements allocated to the AI/ML constituent, including ODD and AI/ML constituent performance monitoring (to support related objectives in Section C.4.1.4.2), detection of OoD input data and data-recording requirements (to support objectives in Section C.3.2.5);
* non-functional requirements allocated to the AI/ML constituent (e.g. performance, scalability, reliability, resilience, etc.); and
* interface requirements.

**Objective DA-03:** The applicant should describe the system and subsystem architecture, to serve as reference for related safety (support) assessment and learning assurance objectives.

**Objective DA-04:** Each of the captured requirements should be validated**.**

**Objective DA-05:** The applicant should document evidence that all derived requirements generated through the learning assurance processes have been provided to the (sub)system processes, including the safety (support) assessment.

**Objective DA-06:** The applicant should document evidence of the validation of the derived requirements, and of the determination of any impact on the safety (support) assessment and (sub)system requirements.

* + 1. **Data Management**

**Objective DM-01:** The applicant should define the set of parameters pertaining to the AI/ML constituent ODD.

**Objective DM-02:** The applicant should capture the DQRs for all data pertaining to the data management process, including but not limited to:

* the data needed to support the intended use;
* the ability to determine the origin of the data;
* the requirements related to the annotation process;
* the format, accuracy and resolution of the data;
* the traceability of the data from their origin to their final operation through the whole pipeline of operations;
* the mechanisms ensuring that the data will not be corrupted while stored or processed,
* the completeness and representativeness of the data sets; and
* the level of independence between the training, validation and test data sets.

**Objective DM-03:** The applicant should capture the requirements on data to be pre-processed and engineered for the inference model in development and for the operations.

**Objective DM-04:** The applicant should ensure the validation to an adequate level of the correctness and completeness of the ML constituent ODD.

**Objective DM-05:** The applicant should ensure the validation of the correctness and completeness of requirements on data to be pre-processed and engineered for the trained and inference model, as well as of the DQRs on data.

**Objective DM-06:** The applicant should identify data sources and collect data in accordance with the defined ODD, while ensuring satisfaction of the defined DQRs, in order to drive the selection of the training, validation and test data sets.

**Objective DM-07:** Once data sources are collected and labelled, the applicant should ensure the high quality of the annotated or labelled data in the data set.

**Objective DM-08:** The applicant should define the data preparation operations to properly address the captured requirements (including DQRs).

**Objective DM-09:** The applicant should define and document pre-processing operations on the collected data in preparation of the training.

**Objective DM-10:** When applicable, the applicant should define and document the transformations to the pre-processed data from the specified input space into features which are effective for the performance of the selected learning algorithm.

**Objective DM-11:** If the learning algorithm is sensitive to the scale of the input data, the applicant should ensure that the data is effective for the stability of the learning process.

**Objective DM-12:** The applicant should distribute the data into three separate and independent data sets which will meet the specified DQRs:

* the training data set and validation data set, used during the model training;
* the test data set used during the learning process verification, and the inference model verification.

**Objective DM-13:** The applicant should ensure validation and verification of the data, as appropriate, all along the data management process so that the data management requirements (including the DQRs) are addressed.

* + 1. **Learning process management**

**Objective LM-01:** The applicant should describe the AI/ML constituents and the model architecture.

**Objective LM-02:** The applicant should capture the requirements pertaining to the learning management and training processes, including but not limited to:

* model family and model selection;
* learning algorithm(s) selection;
* cost/loss function selection describing the link to the performance metrics;
* model bias and variance metrics and acceptable levels;
* model robustness and stability metrics and acceptable levels;
* training environment (hardware and software) identification;
* model parameters initialisation strategy;
* hyper-parameters and parameters identification and setting;
* expected performance with training, validation and test data sets.

**Objective LM-03:** The applicant should document the credit sought from the training environment and qualify the environment accordingly.

**Objective LM-04:** The applicant should provide quantifiable generalisation guarantees.

* + 1. **Training and learning process validation**

**Objective LM-05:** The applicant should document the result of the model training.

**Objective LM-06:** The applicant should document any model optimisation that may affect the model behaviour (e.g. pruning, quantisation) and assess their impact on the model behaviour or performance.

**Objective LM-07:** The applicant should account for the bias-variance trade-off in the model family selection and should provide evidence of the reproducibility of the training process.

**Objective LM-08:** The applicant should ensure that the estimated bias and variance of the selected model meet the associated learning process management requirements.

* + 1. **Learning process verification**

**Objective LM-09:** The applicant should perform an evaluation of the performance of the trained model based on the test data set and document the result of the model verification.

**Objective LM-10:** The applicant should perform a requirements-based verification of the trained model behaviour and document the coverage of the AI/ML constituent requirements by verification methods.

**Objective LM-11:** The applicant should provide an analysis on the stability of the learning algorithms.

**Objective LM-12:** The applicant should perform and document the verification of the stability of the trained model.

**Objective LM-13:** The applicant should perform and document the verification of the robustness of the trained model in adverse conditions.

**Objective LM-14:** The applicant should verify the anticipated generalisation bounds using the test data set.

* + 1. **Model Implementation**

**Objective IMP-01:** The applicant should capture the requirements pertaining to the implementation process.

**Objective IMP-02:** Any post-training model transformation (conversion, optimisation) should be identified and validated for its impact on the model behaviour and performance, and the environment (i.e. software tools and hardware) necessary to perform model transformation should be identified.

**Objective IMP-03:** The applicant should plan and execute appropriate development assurance processes to develop the inference model into software and/or hardware items.

* + 1. **Interference Model verification and integration**

**Objective IMP-04:** The applicant should verify that any transformation (conversion, optimisation, inference model development) performed during the trained model implementation step has not adversely altered the defined model properties.

**Objective IMP-05:** The differences between the software and hardware of the platform used for training and those used for the inference model verification should be identified and assessed for their possible impact on the inference model behaviour and performance.

**Objective IMP-06:** The applicant should perform an evaluation of the performance of the inference model based on the test data set and document the result of the model verification.

**Objective IMP-07:** The applicant should perform and document the verification of the stability of the inference model.

**Objective IMP-08:** The applicant should perform and document the verification of the robustness of the inference model in adverse conditions.

**Objective IMP-09:** The applicant should perform a requirements-based verification of the inference model behaviour when integrated into the AI/ML constituent and document the coverage of the AI/ML constituent requirements by verification methods.

* + 1. **Data and learning verification**

**Objective DM-14:** The applicant should perform a data and learning verification step to confirm that the appropriate data sets have been used for the training, validation and verification of the model and that the expected guarantees (generalisation, robustness) on the model have been reached.

* + 1. **Verification of (sub)system requirements allocated to the AI/ML constituent**

**Objective DA-07:** Each of the captured (sub)system requirements allocated to the AI/ML constituent should be verified**.**

* + 1. **Configuration Management**

**Objective CM-01:** The applicant should apply all configuration management principles to the AI/ML constituent life-cycle data, including but not limited to:

* identification of configuration items;
* versioning;
* baselining;
* change control;
* reproducibility;
* problem reporting;
* archiving and retrieval, and retention period.
  + 1. **Quality and process assurance**

**Objective QA-01:** The applicant should ensure that quality/process assurance principles are applied to the development of the AI-based system, with the required independence level.

* 1. **Development & post-ops AI explainability**

**Objective EXP-01:** The applicant should identify the list of stakeholders, other than end users, that need explainability of the AI-based system at any stage of its life cycle, together with their roles, their responsibilities and their expected expertise (including assumptions made on the level of training, qualification and skills).

**Objective EXP-02:** For each of these stakeholders (or groups of stakeholders), the applicant should characterise the need for explainability to be provided, which is necessary to support the development and learning assurance processes.

**Objective EXP-03:** The applicant should identify and document the methods at AI/ML item and/or output level satisfying the specified AI explainability needs.

**Objective EXP-04:** The applicant should provide the means to record operational data that is necessary to explain, post operations, the behaviour of the AI-based system and its interactions with the end user.

1. **Human Factors for AI**
   1. **AI operational explainability**
      1. **Target Audience for operational explainability**
      2. **Need for operational explainability**
      3. **Anticipated operational explainability modulation**
      4. **Objectives for operations AI explainability**

**Objective EXP-05:** For each output of the AI-based system relevant to task(s) (per **Objective CO-02**), the applicant should characterise the need for explainability.

**Objective EXP-06:** The applicant should present explanations to the end user in a clear and unambiguous form.

**Objective EXP-07:** The applicant should define relevant explainability so that the receiver of the information can use the explanation to assess the appropriateness of the decision / action as expected.

**Objective EXP-08:** The applicant should define the level of abstraction of the explanations, taking into account the characteristics of the task, the situation, the level of expertise of the end user and the general trust given to the system.

**Objective EXP-09:** Where a customisation capability is available, the end user should be able to customise the level of details provided by the system as part of the explainability.

**Objective EXP-10:** The applicant should define the timing when the explainability will be available to the end user taking into account the time criticality of the situation, the needs of the end user, and the operational impact.

**Objective EXP-11:** The applicant should design the AI-based system so as to enable the end user to get upon request explanation or additional details on the explanation when needed.

**Objective EXP-12:** For each output relevant to the task(s), the applicant should ensure the validity of the specified explanation, based on actual measurements (e.g. monitoring) or on a quantification of the level of uncertainty.

**Objective EXP-13:** The AI-based system should be able to deliver an indication of the degree of reliability of its output as part of the explanation based on actual measurements (e.g. monitoring) or on a quantification of the level of uncertainty.

**Objective EXP-14:** The AI-based system inputs should be monitored to be within the operational boundaries (both in terms of input parameter range and distribution) in which the AI/ML constituent performance is guaranteed, and deviations should be indicated to the relevant users and end users.

**Objective EXP-15:** The AI-based system outputs should be monitored to be within the specified operational performance boundaries, and deviations should be indicated to the relevant users and end users.

**Objective EXP-16:** The training and instructions available for the human end user should include procedures for handling possible outputs of the ODD and performance monitoring.

**Objective EXP-17:** Information concerning unsafe AI-based system operating conditions should be provided to the human end user to enable them to take appropriate corrective action in a timely manner.

* 1. **Human-AI-Teaming**

**Objective HF-01:** The applicant should design the AI-based system with the ability to build its own individual situational awareness.

**Objective HF-02:** The applicant should design the AI-based system with the ability to allow the end user to ask questions and to answer questions from the end user, in order to reinforce the end-user individual situational awareness.

**Objective HF-03:** The applicant should design the AI-based system with the ability to modify its individual situational awareness on end-user request.

**Objective HF-04:** If a decision is taken by the AI-based system, the applicant should design the AI-based system with the ability to request from the end user a cross-check validation.

**Corollary objective HF-04:** The applicant should design the AI-based system with the ability to cross-check and validate a decision made by the end user automatically or on request.

**Objective HF-05:** For complex situations under normal operations, the applicant should design the AI-based system with the ability to identify a suboptimal strategy and propose through argumentation an optimised solution.

**Corollary objective HF-05:** The applicant should design the AI-based system with the ability to accept a proposal rejection, upon request by the end user.

**Objective HF-06:** For complex situations under abnormal operations, the applicant should design the AI-based system with the ability to identify the problem, share the diagnosis including the root cause, the resolution strategy and the anticipated operational consequences.

**Corollary objective HF-06:** The applicant should design the AI-based system with the ability to consider the arguments shared by the end user.

**Objective HF-07:** The applicant should design the AI-based system with the ability to detect poor decision-making by the end user in a time-critical situation.

**Objective HF-08:** The applicant should design the AI-based system with the ability to take the appropriate action outside of a collaboration scheme, in case of detection of poor decision-making by the end user in a time-critical situation.

**Objective HF-09:** The applicant should design the AI-based system with the ability to negotiate, argue, and support its positions.

**Objective HF-10:** The applicant should design the AI-based system with the ability to accept the modification of task allocation / task adjustments (instantaneous/short-term).

* 1. **Modality of interaction and style of interface**

**Objective HF-11:** The applicant should design the AI-based system with the ability to understand through the end-user responses or his or her action that there was a misinterpretation from the end user.

**Objective HF-12:** The applicant should design the AI-based system with the ability to notify the end user that he or she misunderstood the information provided through spoken natural language.

**Objective HF-13:** In case of misinterpretation, the applicant should design the AI-based system with the ability to resolve the misunderstanding through repetition, modification of the modality of interaction and/or with the provision of additional rationale on the initial information.

**Objective HF-14:** The applicant should design the AI-based system with the ability to make room for dialogue turns by other participants, keeping silent when needed and not hindering the user (while the user is engaged in a complex activity), sticking to information that answers a given question by other participants, etc.)

**Objective HF-15:** If spoken natural language is used, the applicant should design the AI-based system with the ability to provide information regarding the associated AI-based system capabilities and limitations.

**Objective HF-16:** The applicant should design the syntax of the spoken procedural language so that it can be learned easily by the end user.

**Objective HF-17:** The applicant should design the AI-based system with the ability to transition from verbal natural language to verbal procedural language depending on its own perception of the performance of the dialogue, the context of the situation and the characteristics of the task.

**Objective HF-18:** The applicant should design the gesture language syntax so that it is intuitively associated with the command that it is supposed to trigger.

**Objective HF-19:** The applicant should design the AI-based system with the ability to filter the intentional gesture used for language from non-intentional gesture, such as spontaneous gestures that are made to complement verbal spoken language.

**Objective HF-20:** If gesture non-verbal language is used, the applicant should design the AI-based system with the ability to recognise the end-user intention.

**Objective HF-21:** If non-verbal language is used, the applicant should design the AI-based system with the ability to acknowledge the end-user intention with appropriate feedback

**Objective HF-22:** If spoken natural language is used, the applicant should design the AI-based system so that it can be deactivated for the benefit of other modalities in case of degradation of the interaction performance.

**Objective HF-23:** The applicant should design the AI-based system with the ability to combine or adapt the interaction modalities depending on the characteristics of the task, the operational event, and/or the operational environment.

**Objective HF-24:** The applicant should design the AI-based system with the ability to automatically adapt the modality of interactions to the end-user states, the situation, the context and/or the perceived end user’s preferences.

* 1. **Error Management**
     1. **Contribution of AI-based systems to a new typology of human errors**

**Objective HF-25:** The applicant should design the AI-based system to minimise the likelihood of design-related errors made by the end user.

**Objective HF-26:** The applicant should design the AI-based system to minimise the likelihood of design-related errors made by the human-AI teaming.

**Objective HF-27:** The applicant should design the AI-based system to minimise the likelihood of HAIRM-related errors.

* + 1. **How AI-based system will affect the methods of error management**

**Objective HF-28:** The applicant should design the AI-based system to be tolerant to end-user errors.

**Objective HF-29:** The applicant should design the AI-based system so that in case the end user makes an error while interacting with the AI-based system, the opportunities exist to detect the error.

**Objective HF-30:** The applicant should design the AI-based system so that once an error is detected, the AI-based system should provide efficient means to inform the end user and correct the error.

* 1. **Failure Management**

1. **AI safety risk mitigation**
   1. **AI safety risk mitigation concept**
   2. **AI SRM top-level objectives**

**Objective SRM-01:** Once activities associated with all other building blocks are defined, the applicant should determine whether the coverage of the objectives associated with the explainability and learning assurance building blocks is sufficient or whether an additional dedicated layer of protection, called hereafter safety risk mitigation (SRM), would be necessary to mitigate the residual risks to an acceptable level.

**Objective SRM-02:** The applicant should establish SRM means as identified in Objective SRM-01.