

5

Edge AI Functional Requirements

5.1 Definition

The edge AI field has emerged in recent years, and there is no specific standardised definition exclusively for edge AI system functional requirements (FRs) in ISO or IEEE standards. The concept of functional requirements for systems, including those incorporating AI, can be framed within the broader context of requirements engineering and system functionalities described in existing standards. This work aims to apply the principles of systems and software, hardware, AI, and data engineering standards. These requirements would focus on what functions an edge AI system must perform, considering its unique characteristics, such as local data processing, minimal latency, and resource constraints.

In this context, edge AI system functional requirements are specifications of what an edge AI system is supposed to do, describing the functions or tasks that the edge AI system must perform to meet its designated operations or to fulfil the needs of its users. These requirements define the behaviours, operations, and processes an edge AI system must execute to satisfy its intended purposes.

Functional requirements are crucial in edge AI systems, which combine hardware, software, AI components, and data development. They clearly describe what needs to be built and serve as a foundation for the edge AI system verification, validation, testing and benchmarking. They typically include inputs, expected outcomes, data handling, processing logic, and interactions with other systems.

Functional requirements describe what an edge AI system, application, or product must do to fulfil its intended purpose. They detail the edge AI system's specific behaviours, functions, and interactions under various

conditions. Functional requirements guide the development process and ensure the final edge AI system (product, algorithm, model) meets user needs and business objectives.

Defining FRs for edge AI systems presents a challenge to traditional RE practices. Classical FRs describe deterministic system behaviours that can be precisely specified and verified. However, the core of edge AI systems is ML and DL models whose behaviour is inherently probabilistic and emergent from the data on which it was trained. It is therefore impossible to specify its exact internal logic or guarantee its correctness on all possible inputs. In the case of edge AI systems, the FRs need to address the observable, verifiable interactions of the system that surrounds it, while the edge AI model is treated as a component with a well-defined interface that provides probabilistic outputs, and the FRs specify the deterministic logic that consumes these outputs.

The ISO/IEC/IEEE 29148:2018 [56] standard defines functional/performance requirements as those that describe the system or system element functions or tasks to be performed by the system. Performance is an attribute of function. A performance requirement alone is an incomplete requirement. Performance is normally expressed quantitatively. There can be more than one performance requirement associated with a single function, functional requirement, or task.

A key technique for structuring these requirements is functional decomposition. This method involves breaking down a high-level user need or system capability into a hierarchy of smaller, more detailed, and independently testable functions. For a typical edge AI system, this decomposition follows the flow of data through the system. For example, a high-level function like requiring the detection and response to anomalies can be decomposed into a verifiable sequence of lower-level functions. This sequence can include that the system acquires data from a specific sensor at a defined frequency and resolution; the system applies a specified filter and normalisation algorithm to the raw data; the system passes the pre-processed data to the AI model for inference; the system receives a classification and a confidence score from the model; if the classification is an anomaly and the confidence score is above a specified threshold, the system executes a defined action, such as activating an alarm or sending a notification; and the system logs all inputs, model outputs, and system actions to a specified memory location.

While the specific functions are domain-dependent, it is possible to define a set of domain-agnostic functional building blocks that are common to most edge AI systems. These include an intelligent data pipeline for ingestion and preprocessing, a scalable engine for model training or inference, a module for model deployment and management, and a mechanism for monitoring and feedback. Structuring FRs around these verifiable building blocks allows engineers to specify and test the edge AI system's behaviour rigorously, even when the AI component itself is probabilistic. This approach makes the system's overall function testable, even if the edge AI's reasoning is not.

When creating edge AI functional requirements, it is vital to remember that they should be specific, measurable, achievable, relevant, and time-bound (SMART). By following these guidelines, developers can be sure that the functional requirements are precise and will help edge AI development build the right system.

When defining the edge AI functional requirements, the designers must describe the KPIs and quantitative and qualitative target measures that are verifiable and follow the SMART criteria. Several considerations should be followed when describing edge AI functional requirements, such as being clearly and precisely described to avoid ambiguity and misunderstandings, feasible considering the technological, time, and resource constraints of the system, written in a way that they can be verified through testing or demonstration and relevant to the users' needs and align with the overall objectives of the edge AI system being developed.

Using the information provided in [45] functional requirements for edge AI systems are specifying what the item should do, aligning to the functional suitability quality characteristic outlined in ISO/IEC 25010:2023 System and software quality models.

5.2 Edge AI Functional Requirements, Key KPIs and Measures.

The following section provides examples (structured in tables) of edge AI functional requirements, KPIs, and measures that researchers and practitioners can use as references for their projects. The performances are related to AI algorithms, data quality, data collection, sensors, communication, system latency, processing time, accuracy, resolution, optimisation, etc.

5.2.1 AI Algorithms

Table 5.1 AI algorithms (Prediction algorithms)

FR aspect	FR description
Name	AI algorithms
Definition	Prediction algorithms for energy optimisation at deep-edge.
Description	The ability to take advantage of collected data and optimise energy consumption (residential/office) in a specific period.
What is measured	The optimised energy consumption compared with the actual/measured energy consumption.
KPI	Quantitative [%].
Methods of collection/ measurement and verification/validation	Demonstrator evaluation. Accuracy.
Target Value	Accuracy (in percentage [%]) of AI predicted energy consumption versus actual/measured energy consumption. (e.g., AI model accuracy ≥ 90 % for a specified period of time).

Table 5.2 AI algorithms (Pattern recognition, multi-sensor data)

FR aspect	FR description
Name	AI algorithms
Definition	AI pattern recognition learning algorithms based on multi-sensor data input.
Description	Pattern recognition capabilities - AI algorithms identifying lighting and environmental characteristics and process adjustment parameters based on sensor data analysis.
What is measured	Recognise.
KPI	Qualitative.
Methods of collection/ measurement and verification/validation	Measure ambient light intensity, ambient light colour components and air quality. Process adjustments to light and share data with BAS/HVAC. Performance.
Target Value	Improvement (in percentage [%]) of light conditions in a building/room based on developed/implemented AI algorithms in the end of the project compared to the start of the project (e.g., 40 % improvement in AI algorithm performance compared to the beginning of the implementation).

Table 5.3 AI algorithms (Pattern recognition, camera data)

FR aspect	FR description
Name	AI algorithms
Definition	AI pattern recognition learning algorithms based on RGB Camera data input in natural environment.
Description	Pattern recognition capabilities – AI algorithms identifying grapes in a 2D RGB image.
What is measured	Measure the effective estimated patterns count.
KPI	Pattern recognition algorithms.
Methods of collection/ measurement and verification/ validation	Use of edge device on the autonomous system during the relevant growth period. Strict and manageable manual patterns counting. Comparison with AI estimated counting.
Target value	Yield estimation deviation (in percentage [%]) based on pattern recognition algorithms compared with actual yield (e.g., estimate yield with an error estimation less than 15 %).

5.2.2 Perception

Table 5.4 Perception (Quality)

FR aspect	FR description
Name	Improved perception
Definition	Perception is the ability of a robot to perceive and understand its environment
Description	Increase in perception quality in deployment area measured with relevant metric
What is measured	Quality of human action classification
KPI	Mean average-precision (MAP)
Methods of collection/ measurement and verification/validation	Academic benchmark data. Standard generalization tests on out-of-distribution data.
Target Value	Increase in precision compared to state-of-the-art on edge devices. (e.g., +25 increase in mIoU (mean intersection over union) with respect to state-of-the-art on edge devices).

Table 5.5 Perception (Frames per second)

FR aspect	FR description
Name	Improved perception
Definition	Perception is the ability of a robot to perceive and understand its environment.
Description	Allow more capable models to run on edge hardware by reducing the computational load of foundation models.
What is measured	Frames per second on edge device (FPS).
KPI	Frames per second on edge device.
Methods of collection/ measurement and verification/validation	Academic benchmark data. Standard generalization tests on out-of-distribution data.
Target Value	Achieved percentage increase in frames per second (FPS) (e.g., + 50 % increase in FPS).

5.2.3 Object Detection

Table 5.6 Object detection (Precision)

FR aspect	FR description
Name	Object detection precision
Definition	In a vision-based object detection algorithm (e.g. traffic sign detection), the system is expected to positively classify as object only the objects of a certain class). The precision measures the ratio between the number of correctly classified positives and the total number of positive detections.
Description	A high precision indicates that the system produces reliable positive detections.
What is measured	Correctly classified positives / total classified positive.
KPI	Precision.
Methods of collection/ measurement and verification/ validation	Test run with benchmark of images.
Target Value	Precision in percentage [%] (e.g., precision > 95 %).

Table 5.7 Object detection (Recall)

FR aspect	FR description
Name	Object detection recall
Definition	In a vision-based object detection algorithm (e.g. traffic sign detection), the system is expected to detect all objects of a certain class visible in an input image). The recall measures the ratio between the number of correctly classified positives and the total number of objects of the specified class.
Description	A high recall indicates that the system reliably finds all objects of the given class.
What is measured	Correctly classified positives / total actual positive.
KPI	Recall.
Methods of collection/ measurement and verification/validation	Test run with benchmark of images.
Target Value	Recall in percentage [%] (e.g., recall > 90 %).

5.2.4 Wireless Communication

Table 5.8 Wireless communication (Multiprotocol)

FR aspect	FR description
Name	Communication multiprotocol.
Definition	Seamless integration of wireless communication protocols into AI enhanced gateway.
Description	Wireless communication protocols supported and used by the system.
What is measured	Number of compatible wireless protocols used.
KPI	Quantitative [#].
Methods of collection/ measurement and verification/validation	Demonstrator evaluation. Validation of wireless protocols.
Target Value	Number of different wireless protocols supported by the AI enhanced gateway and its system (e.g., at least 3 different wireless protocols).

Table 5.9 Wireless communication (Range)

FR aspect	FR description
Name	Wireless range.
Definition	Wireless coverage area of the communication protocols integrated in the intelligent gateway.
Description	Wireless communication range optimised for indoors use in a typical residential/office building including normal obstacles.
What is measured	The maximum functional distance between the gateway and nodes/sensors.
KPI	Quantitative [m].
Methods of collection/ measurement and verification/validation	Demonstrator evaluation. Validation of wireless transmission link.
Target Value	Wireless transmission range [m] indoors. (E.g., minimum 30 m inside for at least one communication protocol).

Table 5.10 Wireless communication (Performance)

FR aspect	FR description
Name	Wireless connectivity range.
Definition	Wireless connectivity with different protocols for lighting nodes for indoor and outdoor usage, gaining reliable data routes and long range.
Description	Wireless protocols supported by the system.
What is measured	Typical node to node range for indoor and outdoor usage.
KPI	Quantitative [#].
Methods of collection/ measurement and verification/validation	Demonstrator evaluation: MESH protocols used in intelligent lighting solutions. Verification and validation of indoor distances above a defined range (e.g., > 50 m indoors).
Target Value	Better performance than “comparable” commercial standards at defined distances (e.g., perform better than commercial standards like Zigbee).

5.2.5 Real-Time

Table 5.11 Real-time functioning (System latency)

FR aspect	FR description
Name	Real-time functioning
Definition	To enable real-time use and dynamic learning by minimising time and resource requirements.
Description	Near real-time operation is important for overall user experience and usability of the integrated system.
What is measured	Inference time, end-to-end latency of the system.
KPI	System Latency.
Methods of collection/ measurement and verification/validation	Live Demonstration.
Target Value	Low system latency (near real-time operation if required by the functionality) (e.g., dependent on each functionality (2-60 sec)).

Table 5.12 Video processing real time (Frame processing time)

FR aspect	FR description
Name	Video processing real-time
Definition	In a vision-based AI system, input data often comes in form of a video stream with a certain frequency (e.g. 25 frames per second or fps). The system is said to be “video-real-time” if it can process each frame before the next frame arrives in the input.
Description	The video speed induces an available time budget for each frame inversely proportional to the frame rate. The system is said video real time if the processing time of one frame does not exceed the available time budget for each frame.
What is measured	Time necessary to process one frame.
KPI	Proportion of the time budget used to process one frame.
Methods of collection/ measurement and verification/validation	Test run with benchmark of video input.
Target Value	One frame process time versus time budget versus in percentage [%] (e.g., proportion of time budget < 100 %).

5.2.6 Multi-Sensor

Table 5.13 Multi-sensor measurements (Parameters supported)

FR aspect	FR description
Name	Multi-sensor measurement parameters collection.
Definition	Realtime multi-sensor data collection in industrial environment with edge computing capabilities.
Description	Type of different sensor measurement parameters supported by the system, (ambient light colour, Environmental data gas content of air, IR occupancy).
What is measured	Number of different parameters monitored/ collected.
KPI	Quantitative [#].
Methods of collection/ measurement and verification/validation	Demonstrator evaluation: monitor ambient light, occupancy and air quality. Data collection from sensors measuring each different parameter.
Target Value	An increased number of different sensor measurement parameters supported compared to baseline/starting point (e.g., > 2 different types of sensor parameters).

5.2.7 Data Completeness

Table 5.14 Data completeness (Data handling layer)

FR aspect	FR description
Name	Data completeness
Definition	Degree to which the data in the Data Handling Layer is complete.
Description	Limited availability of data sources, as well as flaws in the setup of the data pipeline could lead to missing data entries in the Data Handling Layer. This could, in turn, negatively impact prediction accuracy.
What is measured	Percentage of empty/nan value in the Data Handling Layer.
KPI	Data completeness.
Methods of collection/ measurement and verification/validation	Data analysis.
Target Value	Increased data completeness [%] compared to baseline/starting point (e.g., 85 %).

5.2.8 Accuracy

Table 5.15 Accuracy (Event detection)

FR aspect	FR description
Name	Accuracy
Definition	Event detection accuracy.
Description	Event detection on difficult data lead to failures in detections. We aim at maximizing the number of correct predictions.
What is measured	The accuracy of the events detection and classification.
KPI	Accuracy.
Methods of collection/ measurement and verification/validation	Inference in real-use conditions and careful labelling. Accuracy measurement on various scenes and conditions.
Target Value	Increased accuracy [%] compared to baseline/starting point (e.g., 80 % accuracy).

Table 5.16 Accuracy (Pose estimation)

FR aspect	FR description
Name	Accuracy of pose estimation algorithms
Definition	Vision-based pose estimation algorithms take an image or a video sequence as an input and outputs the camera pose (6DoF, rotation and translation w.r.t. to a global coordinate frame). The pose estimated as an output should be as near as possible to the ground truth one.
Description	Pose accuracy can be measured in terms of position estimate (deviation from the Euclidean position) and rotation estimation (deviation from the orientation of the camera).
What is measured	Translation deviation in cm, rotation deviation as a quaternion or Euler angles.
KPI	Mean square error in translation over a sequence, Mean angular error in rotation over a sequence.
Methods of collection/ measurement and verification/validation	Test run with benchmark of video input.
Target Value	Decreased error compared to baseline/starting point (e.g., translation error < 5cm, angular error < 5 degrees).

Table 5.17 Accuracy drop (Network reduction)

FR aspect	FR description
Name	Drop of accuracy after network size reduction
Definition	When running AI algorithms on an edge device, the neural networks can be trained on a cloud, then reduced in size to fit in the hardware limitation of the edge device. This reduction can be achieved with quantization and pruning techniques but generally leads to a drop of accuracy compared to the vanilla cloud-trained version.
Description	This requirement defines the acceptable loss of accuracy after the network size has been reduced.
What is measured	Performance (depending on the task: accuracy, precision, recall, similarity etc) in the edge version of the algorithm vs original performance.
KPI	Performance drop.
Methods of collection/ measurement and verification/validation	Test run with benchmark of image input.
Target Value	Acceptable performance drop [%] versus network reduction (e.g., acceptable performance drop of 5 %).

5.2.9 Resolution

Table 5.18 Resolution (Images)

FR aspect	FR description
Name	Resolution of input images
Definition	Edge AI algorithms have limited resources and are able to process images only up to a certain size or resolution. In the case where the input resolution is variable, this functional requirement specifies what is the maximum resolution that the system should be able to process.
Description	Highest image resolution that the system can process.
What is measured	Correct behaviour of the system with a certain resolution image.
KPI	Max image size.
Methods of collection/ measurement and verification/validation	Test run with benchmark of image input.
Target Value	Increased maximum image resolution while maintain correct behaviour compared to baseline/starting point (e.g., size up to 1024 x 1024 pixels).

5.2.10 Optimisation

Table 5.19 Optimisation (Various neural networks)

FR aspect	FR description
Name	Extensibility of the optimization to various types of Neural Networks
Definition	The optimization shall be extensible to allow the optimization to be applicable to various types of Neural Networks.
Description	The optimization algorithm shall be extensible to support various types of Neural Networks, including: CNNs, LSTMs and Transformer models.
What is measured	The ability of the optimization algorithm to optimise various types of Neural Network architectures.
KPI	Achieved / Not achieved [Qualitative].
Methods of collection/ measurement and verification/validation	Optimization runs applied to the specified types of Neural Networks. Results showing optimised models for the above specified types of Neural Networks.
Target Value	Shown applicability of the optimization to the above specified types of Neural Networks.

Table 5.20 Optimisation (Neural networks and secondary hardware)

FR aspect	FR description
Name	Multi-objective Optimization
Definition	The method/techniques shall concurrently optimise a NN for a reference application to improve the accuracy of predictions while improving a secondary HW-related metric on the target platform.
Description	The multi-objective optimization shall automatically design a CNN for a reference application to achieve high accuracy and low latency on the target inferencing platform.
What is measured	Model task performance and secondary HW-related metric.
KPI	[%] Task Performance and Secondary Metric.
Methods of collection/ measurement and verification/validation	The task performance of the optimised network will be measured on the dataset used to optimise the NN and benchmarked with comparable SOTA networks. The secondary HW-related metric of the network will be measured on the target platform. Dataset performance and HW-related metric measured on the target platform.
Target Value	A task performance increase and secondary HW-related metric improvement over comparable SOTA networks

Table 5.21 Optimisation objectives (Extendibility)

FR aspect	FR description
Name	Extendibility of the optimization objectives
Definition	The optimization shall be extendable to support the introduction of additional/alternative optimization objectives.
Description	The algorithm shall support the introduction of additional/alternative objectives in addition to accuracy and latency to accommodate diverse customer interests (e.g., memory).
What is measured	Extendibility of the optimization objectives.
KPI	Achieved / Not achieved [Qualitative].
Methods of collection/ measurement and verification/validation	Additional optimization runs will be performed with alternative/additional objectives to prove the extendibility of the algorithm implementation. Model optimised for additional/alternative objectives.
Target Value	Extendible algorithm to accommodate additional/alternative optimization objectives.

Table 5.22 Photometric optimisation (Image generation)

FR aspect	FR description
Name	Similarity of generated images
Definition	Machine-learning based image generation algorithms such as e.g. face reenactment produces synthetic images that aim to look like realistic versions of the scene. The similarity measures indicate the photometric similarity between a produced images and the ground truth.
Description	When ground truth is available for test benchmark, the photometric similarity can be measured using mean-square-error (MSE) or peak signal-to-noise ratio (PSNR) between images.
What is measured	The MSE represents the cumulative squared error between the generated and the original image, whereas PSNR represents a measure of the peak error. The lower the value of MSE, the lower the error.
KPI	Mean square error or peak signal-to-noise ration.
Methods of collection/ measurement and verification/validation	Test run with benchmark of image input.
Target Value	Increased PSNR [dB] compared to baseline/starting point (e.g., PSNR > 30dB).