

ITU-T Technical Specification

(10/2026)

FG-EAI D2.1

Requirements for multimodal data acquisition procedures for embodied artificial intelligence systems

CAUTION!
PREPUBLISHED TECHNICAL SPECIFICATION

This prepublication is an unedited version of a recently approved Technical Specification. It will be replaced by the published version after editing. Therefore, there will be differences between this prepublication and the published version.



Technical Specification ITU-T FG-EAI D2.1

Requirements for multimodal data acquisition procedures for embodied artificial intelligence systems

Summary

Technical Specification ITU-T FG-EAI D2.1 specifies requirements for multimodal data acquisition procedures for embodied artificial intelligence (EAI) systems. It defines a general procedural framework covering acquisition purposes, data sources and acquisition modes, acquisition targets and coverage objectives, acquisition preparation, process and event recording, data quality and coverage checks, data output, and feedback for additional data acquisition. It also specifies modality-specific requirements for embodiment descriptions, action and observation records, minimum acquisition and recording requirements for each modality, and declarations on the applicability of simulation and synthetic data to real-world use.

The resulting records support data traceability, decisions on data release and use, and subsequent data processing, annotation, training, evaluation and reuse.

Keywords

Embodied artificial intelligence, multimodal data acquisition.

Note

This is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1.0 of the ITU-T FG-EAI Technical Specification on "*Requirements for multimodal data acquisition procedures for embodied artificial intelligence systems*" approved at the 3rd meeting of the ITU-T Focus Group of Embodied AI for Multimedia Technologies (FG-EAI) held on 13-16 October 2026.

Acknowledgements

This Technical Specification was developed in WG2 of FG-EAI, with inputs from PetroChina (Beijing) Digital Intelligent Research Institute and CAICT as contribution to the ITU Focus Group on Embodied AI for Multimedia Technologies (FG-EAI). The development of this document was coordinated by Kei Kawamura (KDDI) as FG-EAI Working Group 2 (WG2) Chair, Qiongqian Yang (Hangzhou Embodied Intelligence Pilot Base Technology) as WG2 Vice-Chair and Editors. In addition, special thanks to Yuntao Wang (FG-EAI Chair), the FG-EAI management team, and the FG-EAI Secretariat.

Editor:	Jian Yang PetroChina (Beijing) Digital Intelligent Research Institute China	Tel:+86 18500367014 Email: yangjian2020@petrochina.com.cn
Editor:	Bohang Wang PetroChina (Beijing) Digital Intelligent Research Institute China	Tel:+86 18610965512 Email: wang_bohang@petrochina.com.cn
Editor:	Haodong Chen CAICT China	Tel:+86 13141490061 Email: chenhaodong@caict.ac.cn
WG2 Chair:	Kei Kawamura KDDI Corporation Japan	Tel: +81-80-5066-9075 Email: ki-kawamura@kddi.com
WG2 Vice-chair:	Qiongqian Yang Hangzhou Embodied Intelligence Pilot Base Technology Co., Ltd. China	Tel: +86 15116306264 Email: yangqiongqian@jszn.cn

© ITU 2026

Some rights reserved. This publication is available under the Creative Commons Attribution-Non Commercial-Share Alike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

For any uses of this publication that are not included in this licence, please seek permission from ITU by contacting TSBmail@itu.int.

Table of Contents

	Page
1 Scope.....	1
2 References.....	1
3 Definitions	1
3.1 Terms defined elsewhere	1
3.2 Terms defined in this Technical Specification	1
4 Abbreviations and acronyms	1
5 Conventions	1
6 General procedural framework for multimodal data acquisition in EAI systems	2
6.1 General framework	2
6.2 Relationship with data platforms, systems and workflows	2
6.3 Outputs of the acquisition procedure.....	3
7 General requirements for multimodal data acquisition procedures	3
7.1 Acquisition purposes	3
7.2 Data sources and acquisition modes.....	4
7.3 Acquisition preparation and recording configuration.....	5
7.4 Formal acquisition and process recording	6
7.5 Data quality and coverage checking.....	6
7.6 Data output and supplementary acquisition feedback	9
8 EAI-specific acquisition and recording requirements	9
8.1 Embodiment description.....	9
8.2 Action and observation recording	10
8.3 Minimum acquisition and recording requirements for each modality	10
8.4 Simulation-to-reality equivalence and synthetic data declaration.....	11
Appendix I – Example of acquisition purpose description and acquisition plan.....	12
I.1 General	12
I.2 Example of acquisition purpose description items.....	12
I.3 Example of modality configuration items	13
I.4 Example focus areas for different acquisition purpose types.....	13
Appendix II – Example of multimodal data acquisition in petrochemical inspection scenarios	15
II.1 General	15
II.2 Example of acquisition purposes and modes.....	15
II.3 Example of acquisition objects, modality configuration and coverage targets	15
II.4 Example of process records and pre-output checking	16

Technical Specification ITU-T FG-EAI D2.1

Requirements for multimodal data acquisition procedures for embodied artificial intelligence systems

1 Scope

This Technical Specification specifies requirements for multimodal data acquisition procedures for embodied artificial intelligence (EAI) systems, including the acquisition workflow and procedures for quality and coverage checks. It addresses the following subjects:

- a) To provide the procedural framework for multimodal data acquisition in EAI systems;
- b) To provide the requirements for planning acquisition purposes, selecting data sources and acquisition modes, identifying acquisition targets and coverage objectives, and configuring acquisition and recording activities;
- c) To provide the requirements for acquisition batch records, process and event records, quality and coverage checks, data output descriptions, and feedback for additional data acquisition.

2 References

- [ITU-T F.748.66] Recommendation ITU-T F.748.66 (2025), *Requirements and framework for embodied artificial intelligence systems*.
- [ITU-T F.TR-ED] ITU-T F.TR-ED, *Technical requirements for embodied artificial intelligence teleoperation data collection systems in real-world scenarios*.

3 Definitions

3.1 Terms defined elsewhere

This Technical Specification uses the following term defined elsewhere:

3.1.1 embodied artificial intelligence [ITU-T F.748.66]: Artificial intelligence integrated into physical systems that interact autonomously with, and adapt to the physical world.

3.2 Terms defined in this Technical Specification

None.

4 Abbreviations and acronyms

This Technical Specification uses the following abbreviations and acronyms:

- AI Artificial Intelligence
- EAI Embodied Artificial Intelligence

5 Conventions

The following conventions are used in this Technical Specification:

- The keywords "it is critical" indicate a possible requirement which would be necessary to be fulfilled (e.g., by an implementation) and enabled to provide the benefits of the use case.
- The keywords "it is expected" indicate a possible requirement which would be important but not absolutely necessary to be fulfilled (e.g., by an implementation). Thus, this possible requirement would not need to be enabled to provide complete benefits of the use case.

- The keywords "it is of added value" indicate a possible requirement which would be optional to be fulfilled (e.g., by an implementation), without implying any sense of importance regarding its fulfilment. Thus, this possible requirement would not need to be enabled to provide complete benefits of the use case.

6 General procedural framework for multimodal data acquisition in EAI systems

6.1 General framework

Figure 1 shows the general procedural framework for multimodal data acquisition in EAI systems. The procedure applies to multimodal data acquired for use in EAI systems and covers the definition of acquisition purposes; the selection of data sources and acquisition modes; the identification of acquisition targets and coverage objectives; acquisition preparation; formal acquisition and process and event recording; data quality and coverage checks; data output; and feedback for additional data acquisition. The general requirements are specified in clause 7, and the EAI-specific acquisition and recording requirements are specified in clause 8.

This Technical Specification does not prescribe specific acquisition devices, platforms, tools or implementation methods. Different implementation methods may be used, depending on the acquisition purpose, data source, acquisition mode and application scenario.

Requirements for data cleaning, data annotation, dataset construction, model training and model evaluation are outside the scope of this Technical Specification. It is critical that acquisition activities retain the metadata and traceability information needed for subsequent processing, annotation, training, evaluation or reuse.

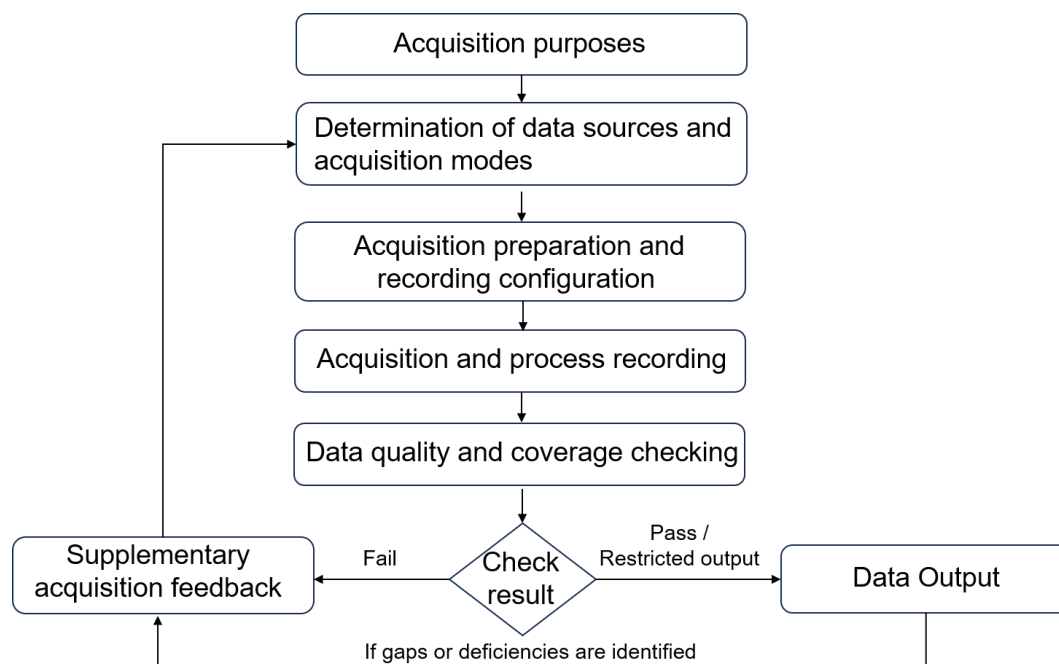


Figure 1 – General procedure framework for multimodal data acquisition in EAI systems

NOTE – Supplementary acquisition feedback may accompany data output or arise during subsequent use.

6.2 Relationship with data platforms, systems and workflows

The multimodal data acquisition procedure links application requirements with downstream data workflows. Application requirements determine the acquisition purpose; acquisition systems perform the acquisition activities; and data platforms store, manage and make the acquired data available for subsequent processing, annotation, dataset construction, training, evaluation or reuse.

The acquisition procedure specified in this Technical Specification covers only the acquisition stage and its direct outputs. The architecture of data platforms, dataset construction systems, model training systems and evaluation systems is outside the scope of this Technical Specification.

It is critical that the outputs of the acquisition procedure include acquisition metadata, process records, check results and information on limitations of use for downstream data workflows.

NOTE – Teleoperation, human demonstration and other human-in-the-loop methods can be used to implement planned acquisition (see clause 7.2.2). The architecture and module requirements of teleoperation data-collection systems, teleoperation-specific data schemas and data-security mechanisms are addressed in [ITU-T F.TR-ED] and are outside the scope of this Technical Specification. This Technical Specification specifies a general procedure that is independent of acquisition mode and data source, together with EAI-specific acquisition and recording requirements that apply, where relevant, across different acquisition modes and data sources.

6.3 Outputs of the acquisition procedure

It is critical that the multimodal data acquisition procedure produce, at a minimum, the following outputs, depending on the acquisition purpose, data source and acquisition mode:

- a) acquisition purpose description;
- b) acquisition plan;
- c) acquisition batch record, including the embodiment description where applicable and modality-specific records specified in clause 8;
- d) process and event records;
- e) quality and coverage check results;
- f) data output description.

Where data insufficiency, insufficient coverage, unclear usage limitation or failure cases are identified, it is critical that supplementary acquisition feedback be produced.

These outputs are used to support traceability, quality checking, coverage checking, data output determination and subsequent data processing, annotation, training, evaluation or reuse.

7 General requirements for multimodal data acquisition procedures

7.1 Acquisition purposes

Before formal acquisition, it is critical that the acquisition purpose be defined. An acquisition activity may serve one or more acquisition purposes. When the same batch of data is intended for multiple purposes, it is critical that the quality criteria, coverage criteria and limitations of use applicable to each purpose be defined separately.

It is expected that acquisition purposes be classified into the following types:

- a) **Pre-training data acquisition**, which is used to support representation learning, general perception, scene and object understanding, and multimodal association capabilities;
- b) **Post-training data acquisition**, which is used to support model behavior optimization, instruction response, preference alignment, learning from feedback and improvement based on failure cases;
- c) **Fine-tuning data acquisition**, which is used to support the adaptation of pre-trained models to specific tasks, skills or application scenarios;
- d) **Evaluation and validation data acquisition**, which is used to support capability evaluation, performance comparison, regression testing, pre-deployment validation or boundary-condition validation;

- e) **Data acquisition for closed-loop improvement**, which is used to collect data from system operation to support failure analysis, model retraining or fine-tuning, and validation of subsequent improvements.

NOTE – The acquisition purposes above can support EAI applications in scenarios such as manufacturing, logistics, household services and industrial inspection. These scenarios can involve dirty, dull or dangerous tasks. Petrochemical inspection, for example, can involve repetitive inspection tasks in dirty or hazardous environments. Appendix II provides an example of applying the multimodal data acquisition procedure to petrochemical inspection scenarios.

7.2 Data sources and acquisition modes

Before formal acquisition, it is critical that the data sources and acquisition modes be selected based on the acquisition purpose, task type, scenario conditions and intended use of the data. It is critical that the data source, acquisition mode, applicable scenarios and limitations of use be recorded.

7.2.1 Data sources

It is expected that data sources be classified into the following types, where applicable:

- a) **Real data**, which refers to data acquired from physical environments, physical objects, embodied systems or real operation processes;
- b) **Simulation and synthetic data**, which refers to data acquired from simulation environments or digital twin environments or using data-generation models.

When the same acquisition activity contains multiple data sources, it is critical that the applicable scope, usage limitation and relationship of each data source be recorded separately.

When simulation-environment data is used as a data source, it is critical that its source, generation condition, ground-truth information, usage limitation and relationship with real-environment data be recorded.

7.2.2 Acquisition modes

It is expected that acquisition modes be classified into the following types:

- a) **Planned acquisition**: an acquisition activity for which the scenario conditions, task objectives, task boundaries, modality configuration and recording rules are defined in advance based on the acquisition purpose, acquisition target and coverage objective. It is expected that planned acquisition be implemented, where applicable, through teleoperation, human demonstration, autonomous execution of specified tasks by a robot, fixed task scripts or simulated task execution. Requirements for teleoperation data-collection systems are specified in [ITU-T F.TR-ED];
- b) **Operation-derived acquisition**: an acquisition activity in which data are generated during autonomous, continuous or event-triggered operation of an EAI system. Where applicable, it is expected that operation-derived acquisition includes autonomous-operation logs, continuous-operation data, event-triggered data, data windows surrounding abnormal events, and data acquired before and after human takeover. Operation-derived acquisition can support closed-loop improvement by collecting data from system operation for failure analysis, model retraining or fine-tuning, and validation of subsequent improvements;
- c) **Simulation and synthetic data acquisition**: an acquisition activity in which data are obtained from simulation environments, digital-twin environments or computational data-generation methods. It is expected that this mode be used for supplementing data relating to rare, dangerous or boundary conditions, failure cases, or broad variations in task or environmental parameters.

When an acquisition activity uses multiple acquisition modes, it is critical that the applicable scope, usage limitation and relationship of each acquisition mode be recorded separately.

7.3 Acquisition preparation and recording configuration

Before formal acquisition, it is critical that the acquisition plan define, at a minimum, the acquisition purpose, data source, acquisition mode, acquisition targets, coverage objectives, embodiment configuration where applicable, modality configuration, acquisition conditions, action and observation recording configuration where applicable, recording rules, temporal-alignment method, spatial-alignment method where applicable, quality-check criteria and coverage-check criteria.

When one acquisition activity corresponds to multiple acquisition purposes, it is critical that the acquisition plan identify the applicable quality checking criteria, coverage checking criteria and usage limitations for each acquisition purpose separately. Appendix I provides examples of acquisition purpose descriptions, modality configurations and purpose-specific focus areas to support the preparation of an acquisition plan.

7.3.1 Acquisition conditions and modality configuration

It is critical that the acquisition plan describe the scenario, task, embodied platform, interaction, environment and safety conditions under which data are generated.

It is critical that the modality configuration identify the modalities necessary for the intended use, identification of the task process, determination of the task outcome, or performance of quality or coverage checks.

It is expected that additional modalities be identified where they provide information useful for data interpretation or subsequent use.

It is of added value to record supplementary modalities beyond those necessary for the intended acquisition purpose where they can support subsequent analysis or reuse.

Modality-specific acquisition and recording requirements are specified in clause 8.

7.3.2 Data format, time synchronization and spatial alignment

It is critical that the data format, measurement unit and encoding method be defined during acquisition preparation. It is expected that a consistent format, unit representation and encoding method be used for data of the same type within an acquisition batch. It is expected that the format version, where applicable, and the data structure, including field definitions and associations among modality data and metadata, be documented. Examples of data formats and organization are provided in Appendix I, clause I.5.

It is critical that all modalities use either a common time reference or documented mappings between their respective time bases. It is expected that the time-synchronization configuration define the allowable synchronization error, acquisition latency, sampling offset and sampling stability based on the acquisition purpose and modality type.

When the acquisition involves spatial information, action trajectories, object positions or coordination among multiple sensors, it is critical that the relevant coordinate relationships or spatial-alignment methods be defined.

7.3.3 Recording rules and checking criteria

It is critical that the recording rules cover process records, event records, abnormal events, human interventions and changes made during acquisition.

For data acquisition for closed-loop improvement, it is critical that the recording rules define the criteria for selecting operational data. It is expected that the selected data include the relevant task or operation context and, where applicable, data windows before and after the event of interest.

It is critical that quality-check and coverage-check criteria be defined before formal acquisition. It is critical that the criteria be consistent with the acquisition purpose, acquisition target, modality configuration and coverage objective.

7.4 Formal acquisition and process recording

During formal acquisition, it is critical that data be collected and recorded in accordance with the acquisition plan and recording rules. It is critical that process records support data traceability, quality and coverage checks, and the description of limitations on data output and use.

7.4.1 Acquisition batch records

It is critical that the acquisition batch record includes, acquisition date or time period, acquisition purpose, data source, acquisition mode, acquisition conditions, modality configuration, time information, spatial information where applicable, process records, event records, acquisition change records and known usage limitations.

Where acquisition involves a physical or simulated embodiment, it is critical that the acquisition batch record be associated with the embodiment description specified in clause 8.1. Where other aspects of the system configuration affect the applicability of the data, it is expected that the acquisition batch record include the relevant software version, computing configuration and acquisition parameters.

For data acquisition for closed-loop improvement, it is critical that the acquisition batch record identify the source system, the reason for data selection and the associated operational records. It is expected that the relevant model version be recorded where it affects data interpretation or subsequent use.

7.4.2 Process records

During acquisition, it is critical that the status of each modality, the task or operational process, key and abnormal events, human interventions and changes to the acquisition configuration be recorded.

For data for which the necessary process records are missing and whose suitability for the intended use cannot be determined, it is critical that such data be excluded from the output for that purpose.

7.4.3 Event records

It is critical that event records be associated with acquisition batch, process records, modality data and time information. It is critical that key events, abnormal events, intervention events and safety-related events be recorded when they affect data quality, coverage or applicability.

For a key event that cannot be associated with time information or a relevant process record, it is critical that the event be excluded from valid event records for quality checking or coverage checking.

7.4.4 Acquisition change records

When the acquisition mode, acquisition conditions, modality configuration, acquisition parameters, time synchronization method, data format or result determination rules change during acquisition, it is critical that the reason for change, change content, change time and affected range be recorded.

When such changes affect the applicability of the data, task-outcome determination or the check criteria, it is critical that the affected data be excluded from the output for the original acquisition purpose unless their suitability for a specified restricted use within that purpose is confirmed in accordance with clause 7.5.3 and the applicable limitations of use are recorded.

7.5 Data quality and coverage checking

Before data output, it is critical that data quality checking and coverage checking be conducted according to the acquisition purpose, acquisition plan, modalities necessary for the intended use, recording rules and coverage targets. It is critical that the check results support the determination of whether the acquired data can be output according to the intended acquisition purpose.

Data quality and coverage checking specified in this clause is limited to determining the output usability of acquired data. The establishment of a complete data quality assessment model, assessment formula or universal threshold requirement is outside the scope of this clause.

7.5.1 Data quality checking

It is critical that data quality checking cover, at a minimum, the following dimensions:

- a) **Record completeness.** The checking is to determine whether the acquisition purpose, data source, acquisition mode, acquisition batch, acquisition conditions, modality configuration, process records, event records and acquisition change records are complete, and whether any missing records affect usage determination;
- b) **Availability of necessary modalities.** The checking is to determine whether all modalities necessary for the intended use are available, whether the minimum recording items specified in clause 8 are present, and whether the modality data are sufficient for the intended use, identification of the task process, determination of the task outcome, analysis of abnormal events and performance of quality checks;
- c) **Temporal and spatial associability.** The checking is to determine whether multimodal data, task process, event records, robot state and environmental state can be associated through time reference, timestamp, coordinate relationship or calibration information where applicable;
- d) **Process and event traceability.** The checking is to determine whether the acquired data can be traced to the acquisition batch, acquisition conditions, task process or operation process, key events, abnormal events, human intervention and acquisition changes;
- e) **Source and mode identifiability.** The checking is to determine whether data from real environments, simulation environments and synthetic generation can be distinguished, and whether data generated under different acquisition modes can be identified;
- f) **Format and unit consistency.** The checking is to determine whether data format, measurement unit and encoding method are consistent for data of the same type within the same acquisition batch, or whether deviations have been recorded and explained;
- g) **Applicability and limitation identifiability.** The checking is to determine whether the applicable scope, known limitations, inapplicable conditions, access conditions and usage constraints of the acquired data are clear.

For fine-tuning data, it is critical that task segments, task boundaries, action processes, state changes and task results be identifiable.

For evaluation and validation data, it is critical that acquisition conditions, input conditions, task boundaries and result determination rules be consistent and traceable. It is critical that evaluation and validation data maintain an identifiable usage boundary from training, post-training and fine-tuning data.

For data acquired for closed-loop improvement, it is critical that the recorded context be sufficient for the intended analysis or subsequent use.

NOTE – A recorded task failure does not in itself indicate a data quality failure. Data recording unsuccessful tasks or abnormal events can satisfy the quality criteria for closed-loop improvement.

7.5.2 Coverage checking

It is critical that coverage checking be conducted according to the acquisition purpose and the coverage targets defined before formal acquisition.

It is critical that coverage checking cover, at a minimum, the following dimensions where applicable:

- a) **Acquisition object coverage.** The checking is to determine whether the defined scenarios, objects, tasks, states, events, data windows or task segments are covered;
- b) **Task or operation process coverage.** For fine-tuning data, evaluation and validation data or operation-derived data, the checking is to determine whether task start, task execution, task

end, result determination, operation trigger, abnormality occurrence and recovery process meet the defined coverage targets;

- c) **State and event coverage.** The checking is to determine whether normal states, abnormal states, boundary states, failure events, human takeover events and safety-related events meet the defined coverage targets;
- d) **Environmental and boundary-condition coverage.** The checking is to determine whether typical environmental conditions, interference conditions and boundary environmental conditions defined before acquisition are covered;
- e) **Data source and acquisition mode coverage.** When an acquisition activity contains multiple data sources or acquisition modes, the checking is to determine whether each data source and acquisition mode is clearly identified and whether it meets the coverage requirements for the intended acquisition purpose.

It is critical that the result of coverage checking describe the covered scope, uncovered scope, reason for insufficient coverage and impact on data use.

7.5.3 Check result classification and decision rules

It is critical that check results be classified as pass, restricted output or fail.

A pass result is assigned when the acquired data satisfy the minimum recording requirements, the availability requirements for modalities necessary for the intended use, traceability requirements, requirements for identifying the data source and acquisition mode, the applicable quality-check criteria, and the coverage objectives defined for the intended purpose.

A restricted-output result is assigned when the data do not fully satisfy the defined requirements but remain suitable for a specified restricted use within the original acquisition purpose. Restricted output applies in the following cases:

- a) Additional modalities identified in the acquisition plan are unavailable and their absence limits the intended use;
- b) Coverage objectives are only partly met;
- c) Temporal or spatial alignment is incomplete but remains sufficient for a specified use;
- d) The data are applicable only to specified scenarios, tasks, objects, conditions or acquisition modes.

For restricted output, it is critical that the data satisfy the minimum recording, necessary-modality, traceability and quality requirements applicable to the specified restricted use.

A failure result is assigned when the data are unsuitable for the original acquisition purpose or their suitability cannot be determined. If the acquisition purpose or data-source information is missing or unverifiable, a failure result is assigned regardless of any proposed restricted use.

For data classified as fail, it is critical that the data be excluded from the output for the original purpose. Before release for a different purpose, it is critical that the data be checked against the quality and coverage criteria applicable to that purpose. If such data are released for a different purpose, it is critical that that purpose, the applicable conditions and limitations of use, and the reasons for failure be recorded.

When quality or coverage checks identify insufficient data, missing records, gaps in coverage, unclear limitations of use or failure cases, it is critical that feedback for additional data acquisition be generated in accordance with clause 7.6.

7.6 Data output and supplementary acquisition feedback

It is critical that data output be based on data quality and coverage check results. It is critical that output data be traceable to the acquisition batch and its associated records and check results, in accordance with clauses 7.4 and 7.5.

It is critical that the data output description include acquisition purpose, data source, acquisition mode, acquisition batch, acquisition conditions, modality configuration, data format, time synchronization and spatial alignment method, quality check result, coverage check result, known limitations and inapplicable conditions. Where applicable, it is expected that licence information and usage constraint information also be included.

Before data acquired for closed-loop improvement are released for model retraining, fine-tuning, or evaluation and validation, it is critical that their suitability be checked against the quality and coverage criteria for the intended use, and that any limitations of use be recorded.

When data quality checking or coverage checking identifies data insufficiency, or when capability gaps, failure cases, abnormal conditions or insufficient boundary-condition coverage are identified during subsequent training, post-training, fine-tuning, evaluation and validation, deployment or operation, it is critical that supplementary acquisition feedback be generated.

It is critical that supplementary acquisition feedback identify the source of the gap, affected acquisition purpose, data to be supplemented, and whether acquisition mode, acquisition conditions, modality configuration or coverage targets need to be adjusted.

After supplementary acquisition is completed, it is critical that process recording, data quality checking and coverage checking be conducted again.

8 EAI-specific acquisition and recording requirements

This clause specifies acquisition and recording requirements specific to embodied artificial intelligence systems that complement the general procedural framework in clause 7. It addresses the coupling among the physical embodiment, its actions and the multimodal observations it produces, and specifies the information needed to interpret and reuse the acquired data.

The design of acquisition systems, the capability and safety requirements of acquisition devices, and the specifications and parameter values of particular data-collection systems, including teleoperation data-collection systems (see [ITU-T F.TR-ED]), are outside the scope of this clause.

8.1 Embodiment description

Where acquisition involves a physical or simulated embodiment, it is critical that each acquisition batch be associated with an embodiment description that enables the acquired data to be interpreted and reused in the context of the system that produced them.

It is critical that the embodiment description record, at a minimum:

- a) the embodiment type and configuration, such as a fixed-base manipulator, mobile manipulator, legged platform or humanoid, and the type of end-effector where applicable;
- b) the kinematic structure, including the number of degrees of freedom, joint types and joint configuration, where needed to interpret spatial or haptic data;
- c) the definition of the embodiment reference frame, including its origin and axis orientation;
- d) the identifier, mounting position and orientation of each acquisition sensor relative to the embodiment reference frame, expressed as extrinsic parameters where applicable.

Where available, it is of added value to provide a machine-readable kinematic model for interpreting spatial relationships among recorded modalities.

When the embodiment configuration or sensor mounting changes during acquisition, it is critical that the change be recorded as an acquisition change record in accordance with clause 7.4.4.

8.2 Action and observation recording

Where action information is necessary for the intended use of data acquired during task execution or system operation, it is critical that the task-level action stream be recorded together with the observation streams.

It is critical that the action representation, measurement unit, value range, update rate and reference frame in which the action is expressed, and the relationship between commanded and executed actions be recorded, where applicable. Where the action reference frame differs from the embodiment reference frame defined in clause 8.1, it is critical that this difference be made explicit. It is expected that the intended action and the observed outcome be recorded, where available.

It is critical that the action stream and the observation streams use either a common time reference or documented mappings between their respective time bases, in accordance with clause 7.3.2, so that each recorded action can be associated with the corresponding observations. It is critical that the method used to align the action and observation streams be recorded.

Where interaction or acquisition latency affects data interpretation, it is critical that the corresponding latency be recorded or be derivable from the recorded timestamps.

8.3 Minimum acquisition and recording requirements for each modality

For each modality necessary for the intended acquisition purpose, it is critical that the acquisition record the minimum descriptive metadata specified in this clause and satisfy the measurement and recording characteristics needed for the acquisition purpose, task-process identification, outcome determination, and quality and coverage checks.

For each recorded quantity, it is critical that the measurement unit be recorded where applicable. Where a sensor requires calibration, it is critical that the calibration method and the most recent calibration date be recorded. Time synchronization among modalities is addressed in clause 7.3.2.

8.3.1 Visual and spatial modalities

For visual and spatial modalities, including image, depth, point-cloud and range-sensing data, it is critical that the acquisition record spatial resolution, frame rate or update rate, intrinsic parameters where applicable, and extrinsic parameters relative to the embodiment reference frame.

For depth, point-cloud and range-sensing data, it is critical that the reference coordinate frame and effective measurement range be recorded. It is expected that synchronized capture across multiple visual sensors be used when the spatial or temporal relationship among the sensors is relevant to the acquisition purpose.

8.3.2 Pose and motion representation

For pose and motion representation, including base or end-effector pose used to interpret visual, spatial or haptic data, it is critical that the acquisition record the state type, whether the value is commanded, measured or estimated, the sampling rate and the reference frame.

It is expected that the sampling rate be sufficient to capture the state changes relevant to the acquisition purpose and the recorded task or operation.

8.3.3 Force and tactile modalities

For force and tactile modalities used for haptic perception and interaction, it is critical that the acquisition record the measured quantity and its dimensionality, such as single-axis force, three-axis force, or six-axis force and torque, together with the measurement range, resolution, sampling rate, sensing location and mounting configuration on the embodiment.

For acquisition purposes involving contact-rich manipulation, it is expected that multi-axis force and torque data at the contact interface be recorded, where available and relevant to the acquisition purpose. It is expected that the contact state be recorded where it affects task-process identification or outcome determination.

8.3.4 Olfactory modalities

For olfactory modalities, limited to odor/smell representation for multisensory interaction, it is critical that the acquisition record the sensor type, data representation, measurement range and resolution where applicable, sampling or update rate, and sensor location.

8.3.5 Audio modalities

For audio modalities, it is critical that the acquisition record the sampling rate and channel configuration. It is also critical that the spatial placement of the audio sensors be recorded when sound-source localization or other spatial audio analysis is within the scope of the acquisition.

8.3.6 Instruction and interaction modalities

For instruction and interaction modalities, including natural-language instructions, gesture inputs, user feedback, teleoperation inputs, human demonstrations and human-intervention events, it is critical that the acquisition record the interaction type, input source, data representation and time information.

It is critical that the interaction data be associated with the relevant task, task phase, system state or executed action. Where an input is processed or transformed, such as through speech transcription or the mapping of gesture input to task-level actions, it is critical that the relationship between the original input and the processed result be recorded.

Where an interaction input constitutes a task-level action stream, it is critical that it be recorded and aligned in accordance with clause 8.2. It is critical that human-intervention events be recorded in accordance with clause 7.4.3.

8.3.7 Other modalities

For other modalities within the scope of this Technical Specification but not listed in this clause, it is critical that the acquisition record the sensor or data-source type, measured quantity, sampling or update rate, and spatial placement relative to the embodiment reference frame where applicable.

8.4 Simulation-to-reality equivalence and synthetic data declaration

For simulation data sources and synthetic-data sources, as identified in clauses 7.2.1 and 7.2.2, it is critical that the acquisition record the simulation or generation configuration needed to assess the applicability of the data to real-world use.

It is critical that a simulated embodiment be described in accordance with clause 8.1; this subclause specifies the additional simulation-specific configuration to be recorded.

It is critical that the configuration record include the physics engine or generation method and its version and, where applicable, the embodiment model and sensor models used, the domain-randomization or parameter-variation ranges applied, the source of ground-truth data, and the known differences from the corresponding real embodiment and environment.

Where simulation data or synthetic data are claimed to be applicable or equivalent to real-world data for a given acquisition purpose, it is critical that the basis for the claim, validation method, applicable scope and known limitations be recorded.

Appendix I

Example of acquisition purpose description and acquisition plan

I.1 General

This appendix provides examples of acquisition purpose descriptions, acquisition-plan items, modality configurations, and data formats and organization that can be considered before formal multimodal data acquisition. These examples are intended to support the implementation of clauses 7.1, 7.2, 7.3 and 8.

The items listed in this appendix illustrate information that can be considered before formal acquisition. The specific items are selected based on the acquisition purpose, data source, acquisition mode, acquisition target and coverage objective.

I.2 Example of acquisition purpose description items

Table I.1 – Example of acquisition purpose description items

Item	Example content	Description
Acquisition purpose identifier	Purpose-001	Unique identifier of the acquisition purpose
Acquisition purpose type	Pre-training, post-training, fine-tuning, evaluation and validation	Used to identify the intended use of acquired data
Acquisition plan identifier	Plan-001	Identifier of the acquisition plan associated with the acquisition activity
Data source	Real-environment data, simulation-environment data, synthetic auxiliary data	Used to identify where data are acquired or generated
Acquisition mode	Planned acquisition, operation-derived acquisition, simulation and synthetic data acquisition	Used to identify how the acquisition activity is organized
Acquisition object	Scenario, object, task, state, event, data window or task segment	Used to define what is expected to be covered by acquisition
Coverage target	Scenario coverage, object coverage, task-process coverage, event coverage or boundary-condition coverage	Used to support coverage checking before data output

Table I.1 – Example of acquisition purpose description items

Item	Example content	Description
Required modalities	Visual, spatial, pose and motion, audio, force and tactile, olfactory or interaction modalities	Used to identify modalities necessary for task identification, result determination or checking
Recording and checking focus	Process records, event records, record completeness, modality availability, temporal and spatial associability, traceability and limitation identification	Used to support traceability, quality checking and coverage checking
Usage limitation	Known limitation, inapplicable condition, access condition or usage constraint	Used to describe the boundary for data output and reuse

I.3 Example of modality configuration items

Table I.2 – Example of modality configuration items

Modality type	Example data	Applicable description
Visual modality	RGB image, video, infrared image, thermal image	Used for scenario recognition, object recognition, state recognition and abnormality identification
Spatial modality	Depth map, point cloud, map, target position, obstacle position	Used for spatial relationship, navigation, obstacle avoidance and scenario reproduction
Pose and motion representation	Trajectory, base or end-effector pose	Used for action process recording, spatial motion representation and association with visual or haptic observations
Force and tactile modality	Contact force, contact torque, tactile array, end-effector pressure	Used for haptic perception in contact operation, grasping, pressing, valve operation and collaborative tasks
Audio modality	Environmental sound, voice instruction, collision sound, alarm sound	Used for interaction recording, abnormality identification and environmental-state determination
Olfactory modality	Odor/smell representation	Used for olfactory perception and multisensory interaction
Interaction modality	Instruction input, user feedback, human takeover, task-level teleoperation input	Used to record human participation, task-level interaction input and interaction state

I.4 Example focus areas for different acquisition purpose types

The following focus areas are illustrative guidance for the acquisition purpose types defined in clause 7.1. They are informative and do not constitute requirements.

Table I.3 – Example focus areas for different acquisition purpose types

Acquisition purpose type	Example focus areas
Pre-training data acquisition	Broad coverage of scenarios, objects, states, interactions and modalities
Post-training data acquisition	System responses, feedback information, correction processes, failure cases and result changes
Fine-tuning data acquisition	The complete execution process of a specific task, including task stages, action processes, spatial poses, object state changes, key events and task results
Evaluation and validation data acquisition	Consistency of acquisition conditions, input conditions, task boundaries, result determination rules, repeatability and separation from training data

I.5 Examples of data formats and organization

Table I.4 provides examples of file formats and data organization approaches that can support acquisition recording and data output. These examples are illustrative and do not require the use of a particular format.

Table I.4 – Examples of data formats and organization

Example	Type	Example application
HDF5	Hierarchical file format	Storing multimodal arrays and associated metadata in groups and datasets.
MCAP	Container format for timestamped messages	Recording sensor messages and operational events with timestamps and message schemas.
RLDS	Organization of sequential interaction data	Organizing task episodes into steps containing observations, actions and associated information.
LeRobotDataset	Dataset format combining data files and metadata	Organizing robot interaction records using Parquet for tabular data, MP4 for video, and metadata linking records to episodes and frames.

The selected format and organization depend on the acquisition purpose, modality characteristics, recording needs and intended downstream use.

Appendix II

Example of multimodal data acquisition in petrochemical inspection scenarios

II.1 General

This appendix provides an example of multimodal data acquisition in petrochemical inspection scenarios. It illustrates how requirements for acquisition purpose definition, acquisition mode determination, acquisition object definition, modality configuration, process recording, pre-output checking and data output, together with the modality-specific requirements in clause 8 such as embodiment description, action and observation recording and visual and spatial data acquisition, can be applied in a specific industry scenario.

This example applies to typical petrochemical inspection objects such as pipe racks, storage tanks, valve groups, pump stations, meters, control cabinets, maintenance platforms, gas detection points and simulated leakage points. Actual acquisition activities can be adjusted according to site safety requirements, device conditions and task boundaries.

II.2 Example of acquisition purposes and modes

Table II.1 – Example of acquisition purposes and modes in petrochemical inspection scenarios

Item	Example content
Acquisition purposes	Fine-tuning data acquisition, evaluation and validation data acquisition, post-training data acquisition and pre-training data acquisition
Planned acquisition	Data acquisition conducted according to preset inspection routes, equipment objects and abnormal conditions
Operation-derived acquisition	Data collected during routine inspection or within data windows before and after alarms, human takeover or task failures, including records selected for closed-loop improvement to address missed detections, false alarms and inspection failures.
Simulation and synthetic data acquisition	Data supplemented through simulation, digital twin or synthetic methods for dangerous, rare, occluded, extreme-lighting boundary conditions

II.3 Example of acquisition objects, modality configuration and coverage targets

Table II.2 – Example of acquisition objects, modality configuration and coverage targets in petrochemical inspection scenarios

Item	Example content
Acquisition objects	Inspection routes, equipment objects, environmental conditions, visually observable leakage-related objects, task processes and events
Coverage targets	Straight routes, turns, narrow passages, platforms, equipment states, viewpoints, distances, abnormal events and intervention events

Table II.2 – Example of acquisition objects, modality configuration and coverage targets in petrochemical inspection scenarios

Item	Example content
Visual and spatial data	RGB image, video, infrared image, thermal image, point cloud, depth map, map, robot pose and target position
Spatial pose and task data	Base or end-effector pose, execution status, inspection route, task phase and task result
Acquisition conditions (metadata)	Temperature, humidity, illumination and noise conditions affecting data acquisition
Interaction and event data	Human instruction, task-level remote interaction input, human takeover, abnormality confirmation, task start, task end, alarm, visible leakage event, failure event and safety stop

II.4 Example of process records and pre-output checking

Table II.3 – Example of process records and pre-output checking in petrochemical inspection scenarios

Item	Example content
Process records	Acquisition batch, acquisition date, acquisition area, acquisition route, acquisition executor, inspection task, task phase, target equipment, task start time and task end time
System and device records	Embodiment description (see clause 8.1), sensor configuration and software version
Temporal relationship	Association among visual data, spatial data, interaction data, and event records.
Spatial relationship	Association among robot pose, simulated visible leakage point location and target equipment location
Event records	Alarm event, abnormal event, human takeover event, safety stop event and acquisition change event
Pre-output checking	Record completeness, required modality availability, visual and spatial data quality, temporal alignment, simulation/synthetic boundary and coverage limitation