SAFEXPL AI N

Safe and Explainable Critical Embedded Systems based on AI

PhDMT0002 Data Collection Log

Version 1.0

Documentation Information

Contract Number	101069595
Project Website	www.safexplain.eu
Contratual Deadline	31.03.2024
Dissemination Level	SEN
Nature	R
Author	Ana Adell
Modified by	Javier Fernández
Reviewed by	Lorea Belategi, Irune Agirre
Approved by	Irune Agirre
Keywords	AI, Functional Safety, Data Management, Data Collection

Table of Contents

1 Review / Modification History

Version	Date	Description Change
V1.0	01/12/2023	First version after complete internal review
V0.3	28/11/2023	Modifications and improvements based on internal review
V0.2	30/08/2023	Modifications and improvements
V0.1	15/05/2023	First draft

2 Objective

The aim of this document is to facilitate the definition of the steps performed in the data collection process of the Data Management phase.

3 Scope

This template applies to the data collection process of the Data Management phase performed through the Artificial Intelligence - Functional Safety Management (AI-FSM).

4 Data Collection

The subsequent subsections encompass all information pertinent to the description of data collection within the project (Data Gathering) and the generation of new data to complete the gathering data (Data Generation).

4.1 Data Gathering

The deliverable generated from this template must include all the information related to the Data Gathering step. This template provides the minimum information that should be collected in this step.

Table 1 collects the information associated with the Data Gathering step.

Table 1. Information related to the Data Gathering step

Data Gathering	
Date	Date of the collection: Format YYYY/MM/DD (Year/month/day)
Responsible	The person who collects the data
Phase of the lifecycle	Data Management
Description	Description of the data collection. It should include information of the data such as: • Format. • Guaranteeing of the data integrity. • Object collected (I.e., people (from kids to elderly), only blonde people, or people from different races).
Data source	Origin of the data, if they have been collected with cameras, sensors, or if it has been obtained from a public dataset (include the link in this case and additional information such as version), etc.
Tools (optional)	Description of the data storage tools employed. Include the required information to replicate their use from scratch.
Data Storage	Include the path to the folder/source where the data is stored.
Observations	Additional information. I.e., specify that it has not been possible to collect the required amount of data to meet the data requirements. Due to this limitation, it is necessary to generate new data.

4.2 Data Generation

The deliverable generated from this template must include the all the information related to the Data Generation step. This template provides the minimum information that should be collected in this step, as outlined in Table 2:

This section collects the information associated with the data generation step.

Table 2. Information related to the Data Generation step

Data Generation			
Date	Date of the collection: Format YYYY/MM/DD (Year/month/day)		
Responsible	The person who generates new data		
Phase of the lifecycle	Data Management		
Description	Description of the data generation process. It has to include the methodology used to generate new data (data augmentation, synthetic data generation, etc.)		
Storage path to source data (optional)	Storage path of the data taken as the source in the generation of new data.		
Storage path to generated data	Include the path to the folder/source where the new data is stored.		
Tools of Data Generation	Tools/programs/frameworks used to generate new data. Include the necessary information for configuration and replicating their use from scratch.		
Description of the Data Generation	Information related to the amount of data generated, how it was generated, etc. It should include enough information to replicate the generation operation.		
Data IDs of Generated Data Traceability among the new data generated from raw or simulation data. It should include the ID of the newly generated data and the identification of the source data file.			
Previous IDs	Previous IDs	**New IDs**	Proposal. Rename the previous identifier by adding the subindex 'GEN_' at the beginning of the name.
Expected results	The set of expected results for data collection or the reason for generating data.		
Observations	Additional information. I.e., problems encountered during the collection.		

5 Acronyms and Abbreviations

Below is a list of acronyms and abbreviations employed in this document:

- AI-FSM – Artificial Intelligence - Functional Safety Management

6 Bibliography

Add here the reference to used bibliography / references (if any).