**3GPP TSG-SA WG6 Meeting #60 S6-241522**

**Changsha, China 15th -19th April 2024 (was S6-241183)**

**Source: Convida Wireless LLC**

**Title: AIML data management**

**Spec: 3GPP TS 23.700-82 V0.3.0**

**Agenda item: 8.3**

**Document for: Approval**

**1. Introduction**

Data operations are part of the learning/inferencing process. Data must be collected and prepared to make it suitable for training/inferencing.

Data collected may not be in the formats suitable for training/inferencing and therefore requires data preparation (e.g. cleaning, scaling, transformation, etc.). Exploratory data analysis provides insights into the statistical properties and relationships of the features of the dataset. If outliers or anomalies are found from EDA requests, then further data preparation may be needed.

With federated learning, it is imperative that exploratory data analyis (EDA) needs to be performed in order for an AIML consumer (e.g. a VAL server) to gain insights into client data without exposing data privacy. Therefore, providing exposure to AIML consumers for performing data operations is an integral part of the AIML support at the application enablement layer.

This contribution proposes a common procedure/ API for data operations targeting data used for training /inferencing. . The data operations considered are: data collection, data preparation, and exploratory data analysis.

**2. Reason for Change**

A data operation procedure which can be used for data collection, data preparation, and exploratory data analysis is proposed..

**3. Proposal**

It is proposed to agree the following changes to 3GPP 23.700-82 V0.3.0.

\*\*\* First change \*\*\*

8.0 Mapping of solutions to key issues

**Table 8.0-1: Mapping of solutions to key issues**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | KI #1 | KI #2 | KI #3 | KI #4 | KI #5 | KI #6 | KI #7 |
| Sol #1 | X |  |  |  |  |  |  |
| Sol #2 |  |  | X |  |  |  |  |
| Sol #3 |  | X |  |  |  |  |  |
| Sol #4 |  | X |  |  |  |  |  |
| Sol #5 |  | X |  |  |  |  |  |
| Sol #6 |  |  | X |  |  |  |  |
| Sol #7 |  |  | X |  |  |  |  |
| Sol #8 |  |  | X |  |  |  |  |
| Sol #9 |  |  | X |  |  |  |  |
| Sol #10 |  |  |  |  |  |  | X |
| Sol #11 | X |  |  |  |  |  |  |
| Sol #12 | X |  |  |  |  | X |  |
| Sol #13 | X |  | X |  |  |  |  |
| Sol #14 |  |  |  |  |  |  | X |
| Sol #15 |  | X |  |  |  |  |  |
| Sol #16 |  |  | X |  |  |  |  |
| Sol #17 |  |  | X |  |  |  |  |
| Sol #18 |  |  |  | X |  |  |  |
| Sol #19 |  |  |  |  | X |  |  |
| Sol #x |  |  | X |  |  |  |  |

\*\*\* Next change – all new text \*\*\*

## 8.x Solution #x: AIML data management procedure

### 8.x.1 Solution description

The following clauses specify procedures, information flows, and APIs for Key Issue #3 to support AIML data management. In this context, the term ”data management” (for AIML) refers to one of the following: data collection, data preparation, and exploratory data analysis.

Data collection is the process of obtaining raw data, whether internally or externally of AIML enablement client, for use in AIML operation. Data preparation is the process of processing raw data into a format suitable for AIML operations (e.g. training, inferencing). Exploratory data analysis (EDA) is the process of analyzing data for general patterns and to obtain statistical properties of the data without having direct access to the data. Since federated learning preserves data privacy, an AIML consumer (e.g. VAL server) utilizes EDA requests to gain an understanding of the data, e.g. to obtain data distributions and data ranges, detect anomalies and outliers, determine relationships between/among features, and other statistical properties.



Figure 8.x.2-1: Procedure for supporting AIML data management.

1. A consumer (e.g. VAL server) makes a request for data management subscription. The request includes a management type indicator (e.g. data collection, data preparation, exploratory data analysis), data management requirements, and a list of AIML enablement clients.

The following is provided for data collection: dataset and data source requirements, sample collection requirements, the time period for which data collection is valid, the location for data collection, etc.

The following is provided for data preparation: a dataset identifier and data preparation requirements.

The following is provided for exploratory data analysis: a dataset identifier, a list of features for statistical analysis, indication for outlier and anomaly reporting, policy for univariate/multivariate analysis, and EDA output formats.

1. The AIML enablement server authorizes the request.
2. The AIML enablement server sends a data management subscription response to the consumer with a status of the request.
3. The AIML enablement server sends data management subscription requests to AIML enablement clients. The request can be for data collection (step 4a), data preparation (step 4b), or exploratory data analysis (step 4c). The request includes the information from the request in step 1.

NOTE: Data collection can reuse A-DCCF procedures

1. The AIML enablement client sends a data management subscription response with a status of accept or reject.
2. The AIML enablement client performs the operations necessary to trigger data management at the UE, e.g. from VAL clients. For data collection, the AIML enablement client performs data collection subscription as described in procedures in TS 23.436. If data is to be collected from the UE, SA4 EVEX mechanism can be reused as described in TS 26.531. For data preparation and exploratory data analysis, the AIML enablement client performs the operation locally.
3. After completing the data management task, the AIML enablement client sends a notification to the AIML enablement server.
4. The AIML enablement server receives and aggregates data management notifications from the AIML enablement clients and sends a notification to the consumer including a status and the requested outputs.

The following is provided for data collection: a dataset identifier, the number of collected samples, the time period and location for which data collection was collected, and a timestamp.

The following is provided for data preparation: the dataset identifier, a report of data preparation outputs, a list of errors detected during data preparation, and a timestamp.

The following is provided for exploratory data analysis: the dataset identifier, statistical output for each feature in the dataset, a report of detected outliers and anomalies, feature correlation information from univariate/multivariate analysis, and a timestamp.

### 8.x.2 Architecture Impacts

This solution does not have architectural impacts.

### 8.x.3 Corresponding APIs

Editor's note: This clause provides the corresponding APIs for supporting the solution.

### 8.x.4 Solution evaluation

The data management solution provides a mechanism for the VAL layer to offload managing and monitoring of AIML data operations to the AIML layer. This solution allows AIML enablement layer to provide value to VAL applications by managing data operations.

.

\*\*\* End of Changes \*\*\*