**3GPP TSG-SA5 Meeting #154 S5-241959d3**

**Changsha, China, 15 - 19 April 2024**

**Source: Intel, NEC**

**Title: New SID Study on AI/ML management phase 2**

**Document for: Discussion**

**Agenda Item: 6.2.1**

**3GPP TSG-SA Meeting #102SP-231780**

Edinburgh, UNITED KINGDOM, 11th Dec 2023 - 15th Dec 2023

**Source: SA WG5**

**Title: New SID: Study on AI/ML management - phase 2**

**Document for: Approval**

**3GPP TSG-SA5 Meeting #152S5-238133**

**Chicago,US, 13-17 November 2023**

**Source: Intel, NEC**

**Title: New SID Study on AI/ML management phase 2**

**Document for: Discussion**

**Agenda Item: 6.2.1**

3GPP™ Work Item Description

Information on Work Items can be found at <http://www.3gpp.org/Work-Items>
See also the [3GPP Working Procedures](http://www.3gpp.org/specifications-groups/working-procedures), article 39 and the TSG Working Methods in [3GPP TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm)

Title: Study on AI/ML management - phase 2

Acronym: FS\_AIML\_MGT\_Ph2

Unique identifier: 1020007

Potential target Release: Rel-19

# 1 Impacts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Affects:** | UICC apps | ME | AN | CN | Others (specify) |
| **Yes** |  |  | X | X |  |
| **No** | X | X |  |  | X |
| **Don't know** |  |  |  |  |  |

# 2 Classification of the Work Item and linked work items

### 2.1 Primary classification

This work item is a …

|  |  |
| --- | --- |
| X | Study  |
|  | Normative – Stage 1 |
|  | Normative – Stage 2 |
|  | Normative – Stage 3 |
|  | Normative – Other\* |

### 2.2 Parent Work Item

|  |
| --- |
| Parent Work / Study Items  |
| Acronym | Working Group | Unique ID | Title (as in 3GPP Work Plan) |
| N/A |  |  |  |

### 2.3 Other related Work Items and dependencies

|  |
| --- |
| Other related Work Items (if any) |
| Unique ID | Title | Nature of relationship |
| 940084 | Study on AI (Artificial Intelligence)/ML (Machine Learning) for Air interface | Rel-18 AI/ML in NG-RAN for Aire interface to be managed |
| 970036 | Support for AI/ML services at application enablement layer | Rel-18 SA6 Work on Application layer support for AI/ML services to be managed |
| 1010005 | New SID for Study on application layer support for AI/ML services | Rel-19 SA6 Study on Application layer support for AI/ML services to be managed |
| 940039 | Study on AI/ML management  | Rel-18 SA5 study on AI/ML management |
| 960037 | Study on Security and Privacy of AI/ML-based Services and Applications in 5G | Rel-18 SA3 study on Security and Privacy of AI/ML-based Services and Applications in 5G |
| 920037 | AI/ML model transfer | Rel-18 SA1 work on AI/ML model transfer in 5GS. |
| 950008 | Study on AI/ML model transfer phase 2 | Rel-19 SA1 study on AI/ML model transfer in 5GS  |
| 1000030 | AI/ML model transfer phase 2 | Rel-19 SA1 work on AI/ML model transfer in 5GS  |
| 980019 | 5GS support for AI/ML-based services | Rel-18 SA2 work on 5GS support for AI/ML-based services |

**Dependency on non-3GPP (draft) specification**: None

# 3 Justification

SA5 has already specified the ML training MnS (Management Service) as part of Rel-17. The group has also studied and is currently progressing to finalise additional aspects of the AI/ML management specifications for Rel-18, focusing specifically on enabling and managing the AI/ML functionalities in the various domains of the 5G system, including the management and orchestration (e.g., MDA defined in TS 28.104), 5GC (e.g., NWDAF defined in TS 23.288) and NG-RAN (e.g., RAN intelligence defined in TS 38.300 and TS 38.401). The AI/ML management capabilities defined in Rel-18 include management and operations for ML training, ML Testing, AI/ML emulation, ML entity deployment and AI/ML inference.

It is also well-know from the latest Rel-19 planning discussions within the 3GPP that almost all the 3GPP WGs, including both the RAN and SA WGs, have already been and plan to continue their engagements in AI/ML relevant features/capabilities standardization development advancements. The OAM standardisation support, led by SA5, continues to play a key role in enabling and facilitating the efficient deployment and operation of the relevant AI/ML features/capabilities.SA1 has specified traffic characteristics and performance requirements for AI/ML model transfer in 5GS in Rel-18, and is studying the phase 2 in Rel-19.

SA2 has defined 5GS support for AI/ML-based services in Rel-18, and is planning the study on architecture enhancement to support the intelligence in 5G Core in Rel-19.

SA3 has studied the Security and Privacy of AI/ML-based Services and Applications in 5G in Rel-18.

SA5 is also studying the MDA (Management Data Analytics) phase 3 in Rel-19.

SA6 is conducting the study to enable support for AI/ML services at application enablement layer in Rel-19.

The management and operation aspects of these AI/ML capabilities need to be investigated in order to identify and address the necessary enhancements to the Rel-18 AI/ML management capabilities.

More and more complex use cases and advanced relevant features in the 5GS can now be facilitated by AI/ML capabilities with support of various type of learning techniques, e.g., but not limited to, Federated Learning, Reinforcement Learning, online training and offline training, and distribute learning across domains (such as E2E smart slicing). This requires more advanced capabilities to enable and manage these learning techniques.

While AI/ML continue to bring a lot of benefits for 5G, the energy consumption aspects associated with AI/ML solutions cannot be neglected. For example, depending on the complexity of the task and structure of the ML model, the energy consumption associated with a specific AI/ML solution/feature could be very different. Some AI/ML solutions may consume relatively higher energy for both training and inference, while some others may consume much more energy for training than inference. Therefore, there is a need to find a way to evaluate and address some means to control the energy consumption associated with AI/ML features/capabilities.

Moreover, based on some high level Rel-18 discussions the trustworthiness (TR 28.908) proves to be an important aspects of AI/ML that need to be further investigated and evaluated, including e.g., the specific trustworthiness indicators, associated measurements and data supporting the quantification of them.

Therefore, a study on AI/ML management phase 2 is needed to address the above-mentioned issues.

# 4 Objective

The objectives of the study on AI/ML management phase 2 include:

WT-1. Continue the study on AI/ML emulation, AI/ML inference coordination and ML knowledge transfer that are leftover from Rel-18.

WT-2. Study the management aspects (LCM CM and PM) of AI/ML functionalities defined by other 3GPP WGs, including

WT-2.1. AI/ML model transfer in 5GS (SA1, Rel-18 WID 920037 and Rel-19 WID 1000030),

WT-2.2. 5GS support for AI/ML-based services (SA2, Rel-18 WID 980019)

WT-2.3. Support for AI/ML services at application enablement layer (SA6, Rel-18 WID 970036),

WT-3. Study the management aspects (LCM CM and PM) of AI/ML functionalities defined by 3GPP SA5 WG, including

WT-3.1. MDA (Management Data Analytics) phase 3 (SA5)

WT-4. Study the AI/ML management and operation capabilities to support different types of AI/ML technologies as needed to support the AI/ML in 5GS, such as Federated Learning, Reinforcement Learning, Online and Offline training, Distributed Learning, and Generative AI.

WT-5. Study the sustainability aspect of AI/ML, including

WT-5.1 Evaluation of energy consumption/efficiency impacts associated with AI/ML solutions for all operational phases (training, emulation, deployment, inference).

WT-6. Further study the trustworthiness aspects related to the AI/ML functionalities in 5GS, including

WT-6.1 Concept of trustworthiness for AI/ML in the context of OAM,

WT-6.2 Data (e.g., measurements, events) to support calculation of trustworthiness indicators.

Note: Whether SA5 can start work on WT-2.1 will be discussed at SA#105 (Sep. 2024) based on the outcome of the related work in the involved RAN WGs(s)

## TU estimates and dependencies

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Work Task ID** | **TU Estimate****(Study)** | **TU Estimate****(Normative)** | **RAN Dependency****(Yes/No/Maybe)**  | **SA Dependency****(Yes/No/Maybe)** | **Non-3GPP Dependency****(EE/ZSM/TMF etc.)** |
| 1 | 1 | 1 | Maybe | Maybe | No |
| 2.1 | 0 | 0 | No | SA1 | No |
| 2.2 | 0.4 | 0.3 | No | SA2 | No |
| 2.3 | 0.4 | 0.3 | No | SA6 | No |
| 3.1 | 0.4 | 0.3 | No | No | No |
| 4 | 1 | 1 | Maybe | Maybe | Maybe |
| 5 | 0.5 | 0.5 | Maybe | Maybe | Maybe |
| 6 | 0.5 | 0.5 | Maybe | Maybe | Maybe |

**Total TU estimates for the study phase: 4.2**

**Total TU estimates for the normative phase: 3.9**

**Total TU estimates: 8.1**

# 5 Expected Output and Time scale

|  |
| --- |
| New specifications {One line per specification. Create/delete lines as needed} |
| Type  | TS/TR number | Title | For info at TSG#  | For approval at TSG# | Rapporteur |
| External TR | 28.9xy | Study on Artificial Intelligence / Machine Learning (AI/ML) Lifecycle Management (LCM) Phase 2 | TSG#104Jun 2024 | TSG#105Sep 2024 | Hassan Al-Kanani (Hassan.Alkanani@EMEA.NEC.COM) |

|  |
| --- |
|  |
|  |  |  |  |
|  |  |  |  |

# 6 Work item Rapporteur(s)

Primary Rapporteur: Yizhi Yao (Intel), yizhi.yao@intel.com

Secondary Rapporteur: Hassan Al-Kanani (NEC), Hassan.Alkanani@EMEA.NEC.COM, TR editor

# 7 Work item leadership

SA5

# 8 Aspects that involve other WGs

Collaboration with SA1 on AIML related requirements, SA2 on 5GC AIML, SA3 on AIML security, SA6 on Application enablement layer AIML and RAN WGs for related requirements.

# 9 Supporting Individual Members

|  |
| --- |
| Supporting IM name |
| Intel |
| NEC |
| Deutsche Telekom |
| US Cellular |
| AT&T |
| Verizon |
| Microsoft |
| China Unicom |
| China Telecom |
| AsiaInfo |
| CATT |
| TELUS |
| ZTE |
| NVIDIA |
| Nokia |
| Telefonica  |
| [P.I. Works](https://piworks.net/) |
| China Mobile |
| Ericsson |
| Telecom Italia |
| Huawei |