3GPP TSG-SA Meeting #94e SP-211328r01

## **Electronic Meeting, 14 ‒ 20 December 2021** (revision of S2-2108158)

## **Source:** OPPO, Samsung

## **Title:** New SID on 5G System Support for AI/ML-based Services

## **Document for:** Approval

## **Agenda Item:** 7.4

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 5G System Support for AI/ML-based Services

**Acronym:** FS\_5GAIML

**Unique identifier:**

**Potential target Release:** Release 18

# 1 Impacts

{For Normative work, identify the anticipated impacts. For a Study, identify the scope of the study}

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

# 2 Classification of the Work Item and linked work items

## 2.1 Primary classification

### This work item is a …

|  |  |
| --- | --- |
|  | Feature |
|  | Building Block |
|  | Work Task |
| X | Study Item |

## 2.2 Parent Work Item

For a brand-new topic, use “N/A” in the table below. Otherwise indicate the parent Work Item.

|  |  |  |  |
| --- | --- | --- | --- |
| Parent Work / Study Items | | | |
| Acronym | Working Group | Unique ID | Title (as in 3GPP Work Plan) |
| AMMT | SA WG1 | 920037 | [AI/ML model transfer in 5GS](https://www.3gpp.org/DynaReport/WiCr--920037.htm) |

### 2.3 Other related Work Items and dependencies

|  |  |  |
| --- | --- | --- |
| Other related Work Items (if any) | | |
| Unique ID | Title | Nature of relationship |
| 860009 | Study on traffic characteristics and performance requirements for AI/ML model transfer in 5GS | SA1 Release 18 study of use cases and potential performance requirements for 5G system support of Artificial Intelligence AI/ML model distribution and transfer (download, upload, updates, etc.), and identification of traffic characteristics of AI/ML model distribution, transfer and training for various applications, e.g. video/speech recognition, robot control, automotive, other verticals |

Dependency on non-3GPP (draft) specification:

# 3 Justification

The AMMT study (AI/ML Model Transfer) in stage-1 has been completed and it is related to how the 5GS supports the transmissions of AI/ML-based services over the application layer. The study addresses use cases and potential performance requirements for 5G system support of application layer Artificial Intelligence (AI)/Machine Learning (ML) model distribution and transfer (download, upload, updates, etc.), and identifies traffic characteristics of AI/ML model distribution, transfer and training for various applications, e.g. video/speech recognition, robot control, automotive, other verticals. Crucially, many of these requirements assume application AI/ML service support at the application client running on the UE, which is currently not specified for the 5GS: while Rel-17 5GS plans to support AI/ML training and inference within the 5GC via NWDAF for network automation purposes, there is no 5GS transport solutions to support for device-based application AI/ML training or inference services. Hence, an evolution of the 5GS is required to provide intelligent transmission support for application AI/ML-based services as proposed in the AMMT study.

The intent of this study is to focus on 5GS architectural and functional extensions so that service providers can leverage 5GS as the intelligent transmission platform to support AI/ML-based services.

# 4 Objective

The intent of this study is to focus on enabling, as defined by SA1 Rel-18 AMMT requirements, TS 22.261, the AI/ML Services & Transmissions with 5GS assistance to support AI/ML model distribution, transfer, training for various applications, e.g. video/speech recognition, robot control, automotive etc. for the following three main types of AI/ML operations:

* AI/ML operation splitting between AI/ML endpoints
* AI/ML model/data distribution and sharing over 5G system
* Distributed/Federated Learning (FL) over 5G system

The following describes the objectives of this study on how the AI/ML service providers could leverage 5GS as the platform to provide the intelligent transmission support for application layer AI/ML operation.

1. Objective 1 (WT#1.1): Study the possible architectural and functional extensions to support the Application layer AI/ML operations defined in TS 22.261, more specifically:
2. Support monitoring of network resource utilization in the 5G system relevant to the UE in order to support Application AI/ML operation with AI/ML model provider.
3. 5GS information exposure extensions for 5GC NF(s) to expose UE and/or network conditions and performance prediction (e.g. location, QoS, load, congestion, etc.) and whether and how to expose such information to the UE and/or to the authorized 3rd party to assist the Application AI/ML operation.
4. Enhancements of external parameter provisioning to 5GC (e.g. expected UE activity behaviors, expected UE mobility, etc.) based on Application AI/ML operation.
5. Investigate the enhancements of BDT and AF influence that could be used to assist the AI/ML operations.

NOTE 1: Any security, privacy and user consent related aspects are expected to be handled by SA3, and if any, to provide guidance for specific SA2 impact.

NOTE 2: Any potential impact towards NWDAF will be coordinated with eNA study in Rel-18.

1. Objective 2 (WT#1.2): Study possible QoS, Policy enhancements to support Application AI/ML operational traffic while supporting regular (non Application-AI/ML) 5GS user traffic.
2. Objective 3 (WT#1.3): Study whether and how 5GS provides assistance to AF and the UE for the AF and UE to manage the FL operation and model distribution/redistribution (i.e. FL members selection, group performance monitoring, adequate network resources allocation and guarantee, .) to facilitate collaborative Application AI/ML based Federated Learning operation between the application clients running on the UEs and the Application Servers.

In Rel-18, all UEs who are participating in the given Application AI/ML operation are served by the same S-NSSAI as the AF does.

In Rel-18, roaming is not supported.

## TU estimates and dependencies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Work Task ID | TU Estimate  (Study) | TU Estimate  (Normative) | RAN Dependency  (Yes/No/Maybe) | Inter Work Tasks Dependency |
| WT#1 | 7.5 TUs | 5.5 TUs | Yes | N/A |
| WT#1.1 | 3.5 TUs | 2.5 TUs | Maybe | WT#1.1 is self-contained |
| WT#1.2 | 1.5 TUs | 1 TUs | Yes | WT#1.2 has dependency on WT#1.1 |
| WT#1.3 | 2 TUs | 1.5 TUs | No | WT#1.3 has dependency on WT#1.1 and WT#1.2 |

Total TU estimates for the study phase: 7.0

Total TU estimates for the normative phase: 5.0

Total TU estimates: 9 + 7 = 12

# 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 |
| TR | 23.xxx | Study on 5G System Support for AI/ML-based Services | TSG#96 (Jun 2022) | TSG#97 (Sept. 2022) | Tricci So, OPPO, [tricci.so@oppo.com](mailto:tricci.so@oppo.com)  David Gutierrez Estevez, Samsung, [d.estevez@samsung.com](mailto:d.estevez@samsung.com) |

|  |  |  |  |
| --- | --- | --- | --- |
| Impacted existing TS/TR {One line per specification. Create/delete lines as needed} | | | |
| TS/TR No. | Description of change | Target completion plenary# | Remarks |
| {e.g. "22.281"} | {Possible values:  - either free text (e.g. “CS aspects to be removed")  - or “Specification to be withdrawn”} | {e.g. "TSG#89"} | {Free text} |

# 6 Work item Rapporteur(s)

Tricci So, OPPO, [tricci.so@oppo.com](mailto:tricci.so@oppo.com) (work planning)

David Gutierrez Estevez, Samsung, [d.estevez@samsung.com](mailto:d.estevez@samsung.com) (TR editing)

# 7 Work item leadership

SA2

# 8 Aspects that involve other WGs

Alignment with relevant requirements from SA1 normative work on traffic characteristics and performance requirements for AI/ML model transfer in 5GS.

Coordination is expected with SA4 in the context of communication between AI/ML model user and provider.

Coordination is expected with SA3 to support security, privacy integrity and user consent issues.

Any Charging and OAM aspects are to be addressed in SA5.

Coordination may be required with SA6 for specific application aspects.

# 9 Supporting Individual Members

|  |
| --- |
| Supporting IM name |
| OPPO |
| Samsung |
| vivo |
| Xiaomi |
| Tencent |
| Toyota |
| Matrixx |
| AT&T |
| Convida Wireless |
| China Unicom |
| SK Telecom |
| KT |
| LG Uplus |
| ETRI |
| NTT DoCoMo |
| China Telecom |
| OnePlus |
| Xidian University |
| Shanghai Jiao tong University |
| Interdigital |
| LG Electronics |
| CAICT |
| DENSO |
| Panasonic |
| Rakuten |
| ZTE |
| Telefonica |
| BT |
| Intel |
| China Mobile |
| Alibaba |
| Fujitsu |
| Charter |
| Inspur |
| Harman |
| SaankhyaLabs |
| Oracle |
| TCL |
| Tejas Networks |
| Zhejiang Lab |
| Verizon |
| ZEKU |
| IIT Bombay |
| Purple Mountain Laboratories |
| Reliance |
| EST |
| DISH Network |
|  |