**3GPP TSG-SA WG4 Meeting #131-bis-eS4-250551\_r02**

**E-meeting, 11 - 17 April 2025**

**Source: Samsung Electronics Co., Ltd**

**Title: [FS\_AI4Media] pCR and discussion on conclusions (TR 26.927)**

**Agenda item: 9.6**

**Document for: Agreement**

**1. Introduction**

This contribution provides an update and discussion on the conclusions for TR 26.927.

**2. Discussion**

The latest text agreed for further work on the conclusion of TR 26.927 is separated into two parts:

1. Text in TR 26.927 and Function PD v1.5.1: identified work for Rel-20
2. Currently only in the Functional PD v1.5.1: identified work for Rel-20 or beyond.

Related to 1), a corresponding WID is drafted in tdoc S4-250554. SA4 should, however, be aware that IMS related work may have SA2 related aspects, and that there is on-going discussion in SA2 regarding a potential WT for AI/ML in IMS in NGRTC\_Ph3. SA2 is expected to make a decision on the final list of NGRTC\_Ph3 WTs during their SA2#168 meeting. As such, the identified aspects in 1) may be better suited as a short study item to identify specific Key Issues (in-line with SA2 aspects), followed by a prompt normative stage 3 work within the same release of Rel-20. It is left to the interested parties in SA4 to decide on the best approach to proceed with the identified aspects in 1).

Related to 2), the text in the Function PD v1.5.1 is updated and proposed as a pCR to be included into the next version of TR 26.927 v1.0.0.

If a study related to 2) is introduced in Rel-20, for example to continue the evaluations in the current study, and if it is decided that 1) should also start from a study, still there should only be one SA4 AI/ML study in Rel-20, but with a clear indication of the separation of the different tracks for the work identified in 1) and 2).

**3. Proposal**

It is proposed to agree the following changes to 3GPP TR 26.927 v1.0.0. New changes against the text in the PD are highlighted in yellow.

\* \* \* First Change \* \* \*

# 7 Conclusion

AI/ML in media services involve the use of AI/ML models to perform media processing, typically with video or audio media as the input into an AI/ML model, resulting in an output which may be a version of processed video or audio media (e.g. picture enhancement, audio translation), a specific description of the input media itself (e.g. labelling in object recognition) or a completely new media (e.g. sign/text translation to speech or video). In order to support such AI/ML based media processing, three scenarios have been documented:

- UE device AI inferencing

- AI inferencing in the network

- Split AI inferencing between the UE and the network.

In this study, the broad findings for AI/ML model transfer in TR 22.874 [2] have been further analysed with specific focus on media-based AI/ML use cases and scenarios. In particular this document describes how AI/ML models and data may be distributed over the 5G system and documents the split AI/ML operations between different AI/ML endpoints (noticeably the UE and the network), and the compression of AI/ML model data and intermediate data. Due to the broad range of applications for AI/ML based media processing, as well as the wide diversity of different AI/ML models available for same application, feasibly evaluations for a given set of scenarios are documented in TR 26.847 [20] as part of this study.

Based on the core use cases, functional architectures are presented for basic AI/ML model distribution, split AI/ML operation and distributed/federated learning. Different AI user plane data components have been identified and documented (AI model data, intermediate data, inference input and output data), and a set of logical AI functions have been defined.

The identified logical AI functions are further mapped to the 5G system, addressing the underlying 5GMS/RTC and IMS DC architectures. The mapping of such AI media use cases to the different architectures and their relevant procedures describes the provisioning, capability discovery/negotiation and delivery session support for the delivery of AI data components and the use of required AI media functions at different endpoints according to the service configuration negotiated. Architecture variants for three different collaboration scenarios are also introduced, each with a different level of MNO network support for AI/ML functions.

Based on the details in the report, the following next steps are identified:

Normative work in release-20:

For collaboration scenario 3 IMS services:

- Recommend stage 3 normative work on the support of AI/ML model distribution and operation in IMS.

- Define the support of the configuration, delivery, compression, and processing of AI/ML data in a new specification.

- Extend TS 26.114 and TS 26.264 specifications to support AI/ML data delivery in IMS services, as identified in clause 5.4.

- Extend TS 26.114 and TS 26.264 specifications to support AI/ML media processing in IMS services, as identified in clause 5.5.

- Specify support for AI/ML data signalling and negotiation, including support for split.

- Select interoperable formats for AI/ML model data and intermediate data.

- Collaborate with SA2 on related matter where necessary.

New Study in Release 20 or beyond:

For collaboration scenarios 1 (OTT) and 2 (Hosting):

- Further study, identify and document the traffic characteristics of the AI/ML data components (as defined in clause 5.3.1 and detailed in clause 6) for the relevant use cases, as introduced in TR 26.847.

- Further study and identify any potential needs for new QoS identifiers, metrics and/or QoS procedures to support the delivery of the AI/ML data components based on the architectures in TS 26.501, TS 26.506, and TS 26.114 for 5GMS, RTC, and IMS respectively.

For collaboration scenario 3 non-IMS services:

- Further study and investigate stage 2 aspects for the architectures in TS 26.501 (5GMS) and TS 26.506 (RTC), identifying potential key issues related to:

- The support of AI/ML model distribution and operation, based on details in clause 5.3.6.

- The support of split AI/ML inferencing between the UE and the network, based on details in clause 5.3.5.

- The support of distributed/federated learning, in particular SA2 defined features, as identified in clause 5.3.7.

Collaborate with SA2 on related matter where necessary for Release 20 or beyond.

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