**Agenda item:** 9.7

**Source:** vivo

**Title:** Use Case on Split Model Update

**Document for** Discussion andAgreement

1. Introduction

In this contribution, we provide a new Use Case on Split Model Update which has also mentioned in SA2 TR 22.874.

1. Proposed changes

--------------------------------------------- Start of Change ----------------------------------------------------------------------------

## 4.x Split Model Adjustment during ongoing AI/ML service

Based on clause 5.5 of TR 22.874 [1], this use case covers all the cases where when the AI/ML models are computing intensive, the work tasks can be fully or partially offloaded to the network and the AI/ML split points can be dynamically adjusted by considering various factors such as UE capabilities (e.g. processing capability/computation resources), service performance, intermediate data volume, and network conditions such as bandwidth etc.

The AI/ML models are set to have different candidate split points and each candidate split point has different workload and communication requirements, as well as intermediate data characteristics. A policy decision point for the media task will adjust the split points of the AI/ML model for an ongoing service based on the factors of current UE’s capabilities, communication performance, intermediate data volume, network conditions etc. to make sure that the media work task can be executed well, guaranteeing the UE experience and avoiding service interruption.

For the 5G media system, both UE capabilities and network conditions are required to be monitored and used as some of the considering factors when updating the AI/ML model split points for an ongoing service; the UE and network can then inference based on the newly updated AI/ML split models in real time.

--------------------------------------------- End of Change ---------------------------------------------------------------------------

1. Proposal

# We propose to update clause 4 of the permanent document with the above proposed changes.