**3GPP TSG RAN WG2 Meeting #131 R2-250xxxx
Bengaluru, India, August 25 – 29, 2025**

**Title: [Draft] Reply LS on AI/ML functionality activation**

**Response to: R2-2505045 (R4-2508085)**

**Release: Release 19**

**Work Item: NR\_AIML\_air-Core**

**Source: Apple [to be RAN2]**

**To: RAN4**

**Cc: RAN1**

**Contact person: Peng Cheng**

 **pcheng24@apple.com**

**Send any reply LS to: 3GPP Liaisons Coordinator,** **mailto:3GPPLiaison@etsi.org**

**Attachments: None**

1 Overall description

RAN2 thanks RAN4 for the LS on AI/ML functionality activation. RAN2 discussed the two interpretations provided by RAN4 and concluded that it is necessary to clarify the following RAN2 understanding:

**Agreements**

1. RRC processing delay shouldn’t be impacted by the model loading delay
2. If the UE is ready for inference (i.e. the model is ready to generate inference result based on configuration) by end of RRC processing delay, it reports configuration applicable. If not, it reports configuration inapplicable and doesn’t set the release flag. The network is not expected to release inference configuration.
3. Once the configuration is applicable (i.e. the model is ready to generate inference result based on configuration), UE reports applicability to network via UAI (applicable to all AI/ML based CSI reporting).
4.
5. On the time duration for an AI functionality to become available for inference, RAN2 conclude that it is up to UE implementation from RAN2 point of view and no further RAN2 work.
6. RRCReconfigurationComplete containing applicability reports has a processing latency requirement of 16 ms with respect to the reception of RRCReconfiguration, from RAN2 point of view.

2 Actions

**To RAN4**

**ACTION:** RAN2 respectfully requests RAN4 to take above RAN2 agreements into account.

3 Dates of next TSG RAN WG2 meetings

TSG-RAN WG2 Meeting #131-bis October 13 to 17, 2025 Prague, CZ

TSG-RAN WG2 Meeting #132 November 17 to 21, 2025 Dallas, US