During the AI-ML main session, a discussion took place on confirming the following tentative agreement from RAN4#112bis:

* **Option 1**: Conduct the ~~conformance~~ testing for new/updated AI model/functionality before its deployment in already deployed UEs (considering the UE hardware that it will be used in)
  + FFS on the feasibility
* **Option 2**: Design the test to verify the performance monitoring and proactive recovery from potential performance degradation
  + Depend on the other WG progress
  + Monitoring can be used for managing fallback, model update/model switching/model transfer, if applicable
* **Option 3**: Capture model input during conformance testing for later testing of new models.
* Other options are not precluded.

Following the discussion, the following text is proposed to capture the discussion:

When operating in the field, two aspects of UE operation may impact performance:

* Update or fine tuning of models, which may impact model performance
* Data draft / mismatch between the conditions encountered by the UE in the field and the training data for the model, which may impact model performance.

For dealing with update or fine tuning of models, two options may be available. The options are not mutually exclusive:

* **Option 1**: Conduct the validation for new/updated AI model/functionality before its deployment/activation in already deployed UEs
  + Validation takes into account the UE hardware in which the model is to be deployed/activated.
  + FFS on the feasibility
  + One possibility for consideration for option 1 is to capture model input (and if needed other test data such as ground truth) during conformance testing. This stored data can later be used to validate new or updated models.
  + Other options not precluded.
* **Option 2**: Using performance monitoring and LCM procedures
  + Performance monitoring will be designed in other groups
  + RAN4 may consider the need and feasibility of requirements and tests to ensure consistency and accuracy of monitoring metrics or other monitoring related data sent from the UE, and set requirements as feasible/needed.
  + Monitoring can be used for managing fallback, model update/model switching/model transfer, if applicable

For dealing with draft / mismatch between the conditions encountered by the UE in the field and the training data for the model, which may impact model performance, monitoring is needed.

* + Performance monitoring will be designed in other groups
  + RAN4 may consider the need and feasibility of requirements and tests to ensure consistency and accuracy of monitoring metrics or other monitoring related data sent from the UE, and set requirements as feasible/needed.
  + Monitoring can be used for managing fallback, model update/model switching/model transfer, if applicable