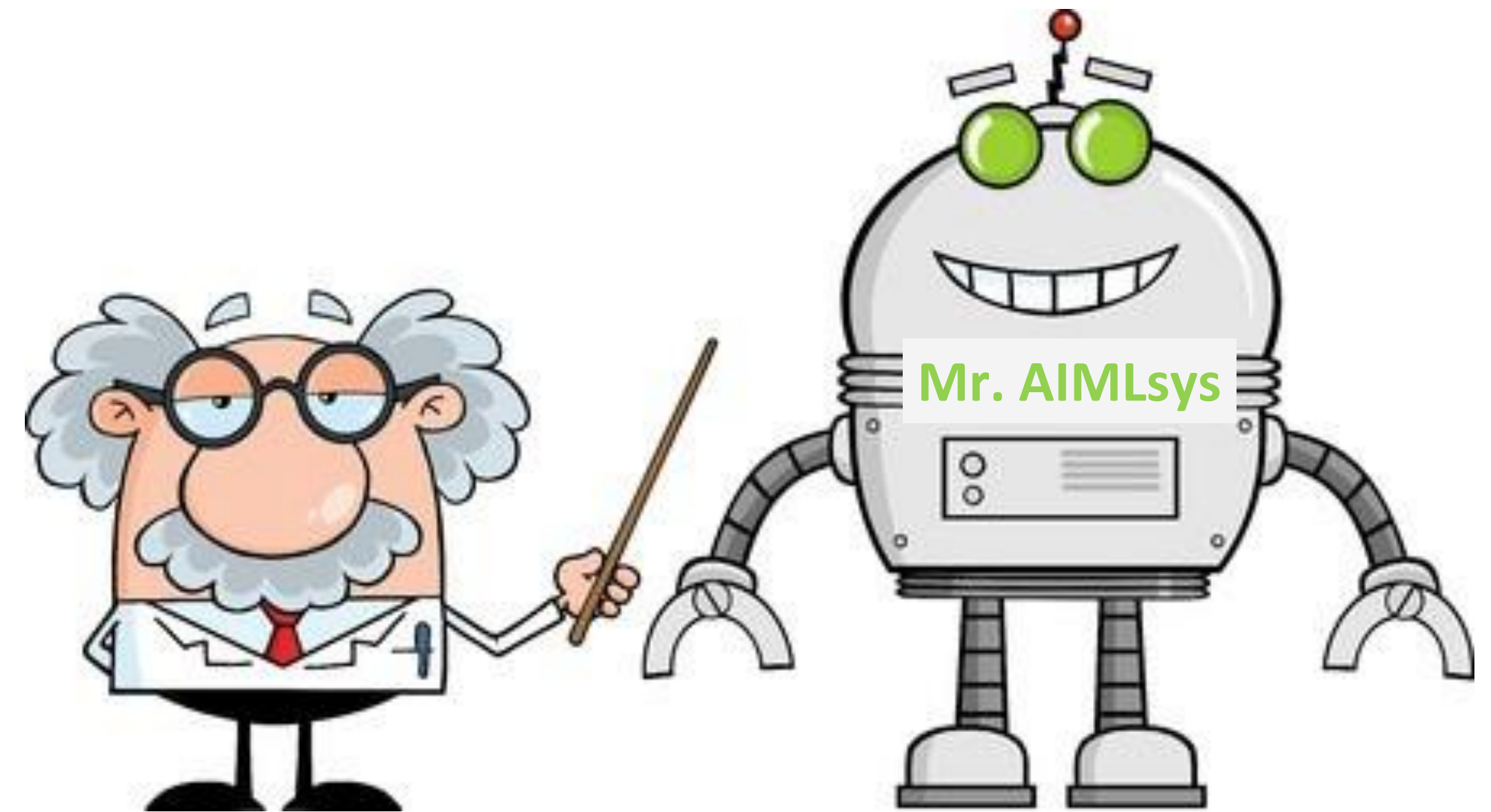
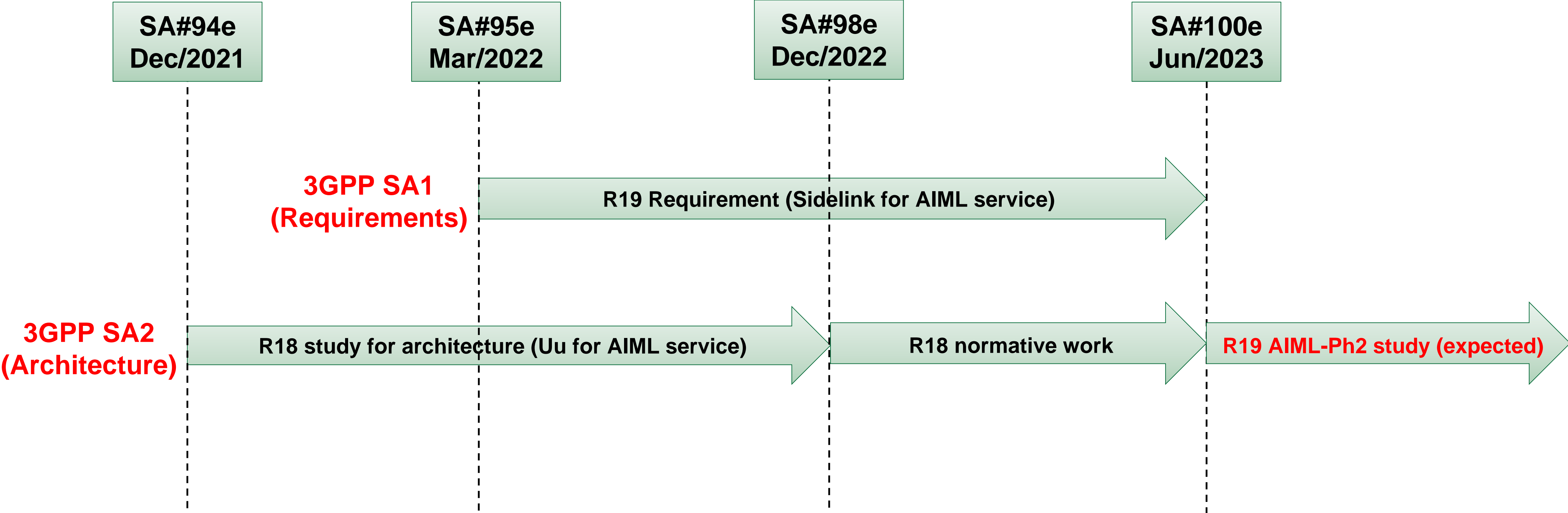


R19 AIML-Phase2 work for 5G architecture



Timeline for R19 AIML-Ph2



Summary

- Leveraging Sidelink for distributed learning/inference (new SA1 requirements)
- Information exposure to UE
- AIML-Translation between application request and 5GC APIs
- 5GS assisted Cross domain AI training and inference (Vertical-FL)
- 5GS assisted Transfer Learning

Leveraging Sidelink for distributed learning/inference

- **Scenario-1: Sidelink based task offloading**

Offloading AI inference task from one UE to another UE in proximity.

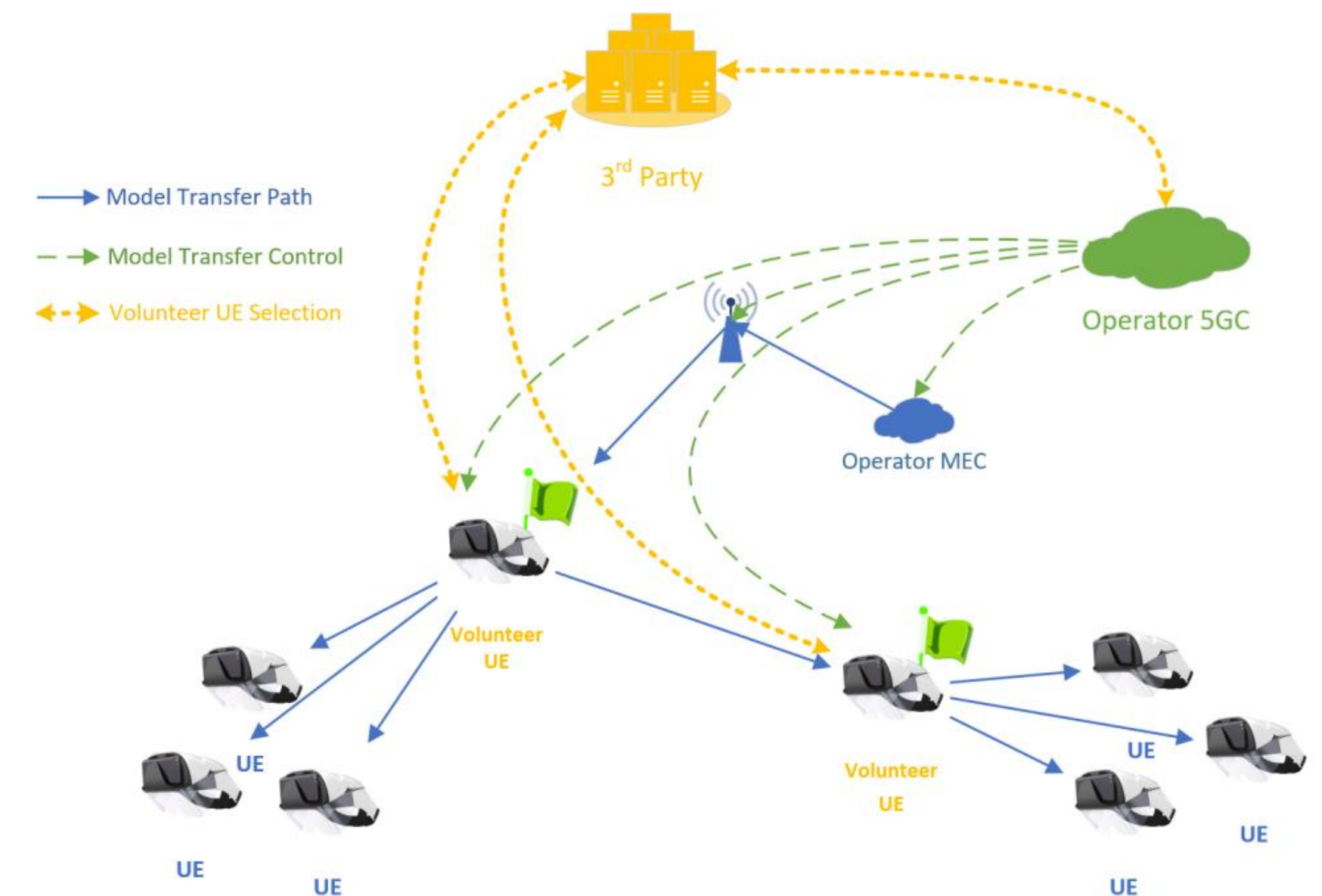
By doing so, the UE who has interesting data set can be used while the computation work task can be offloaded to others, and Uu interface impact can be minimized or avoided.

- **Scenario-2: Sidelink based model distribution**

Model can be shared by volunteer UE and save resource from Uu interface. The model distribution can be finished in time when QoS in Uu interface cannot be fulfilled

- **Scenario-3: Sidelink based federated/distributed learning**

It can involve the UEs who are out of coverage or has bad QoS in Uu interface for a federated learning. For example, when UE moves to a tunnel the Federated Learning can still be performed without losing the UE.



■ Gap analysis

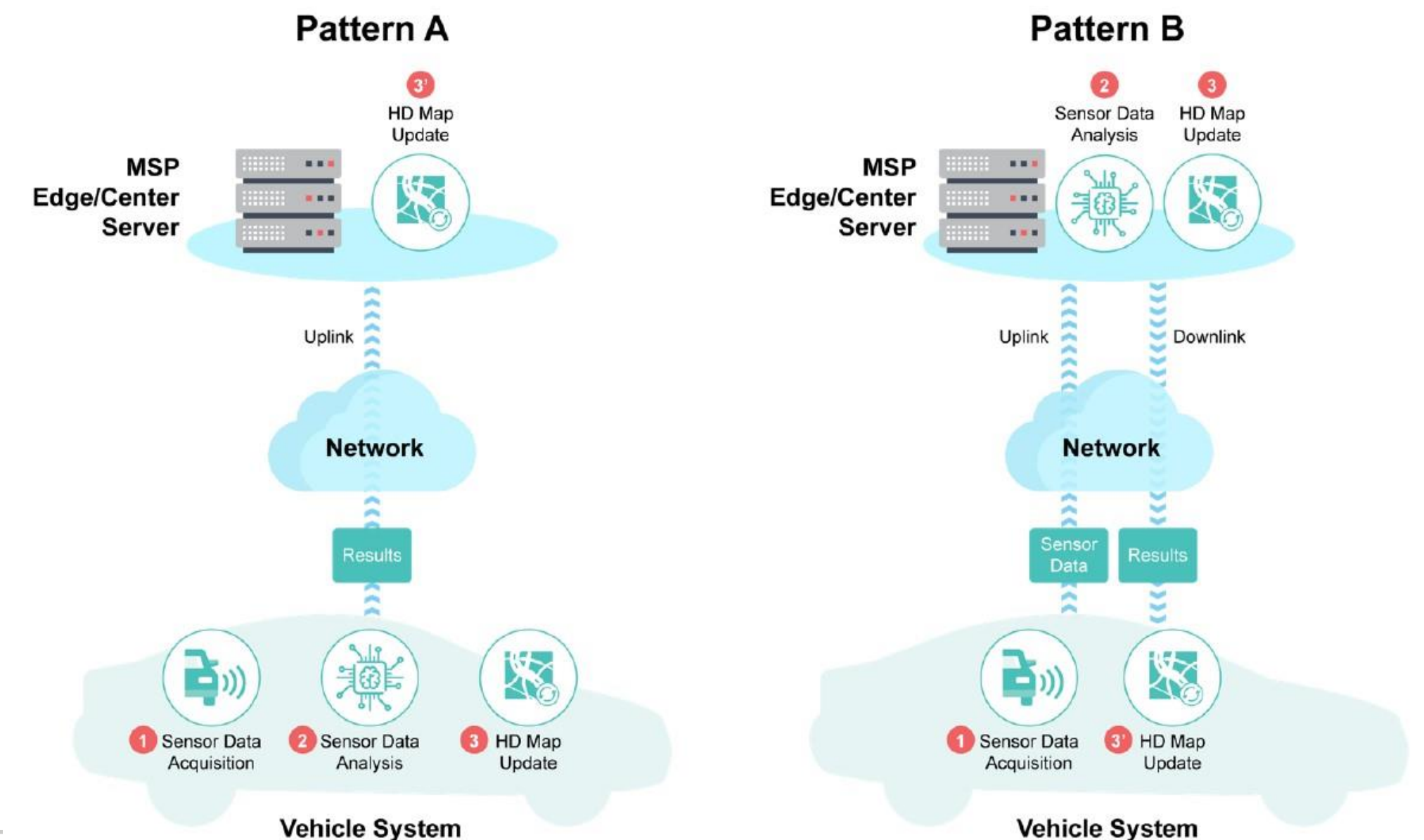
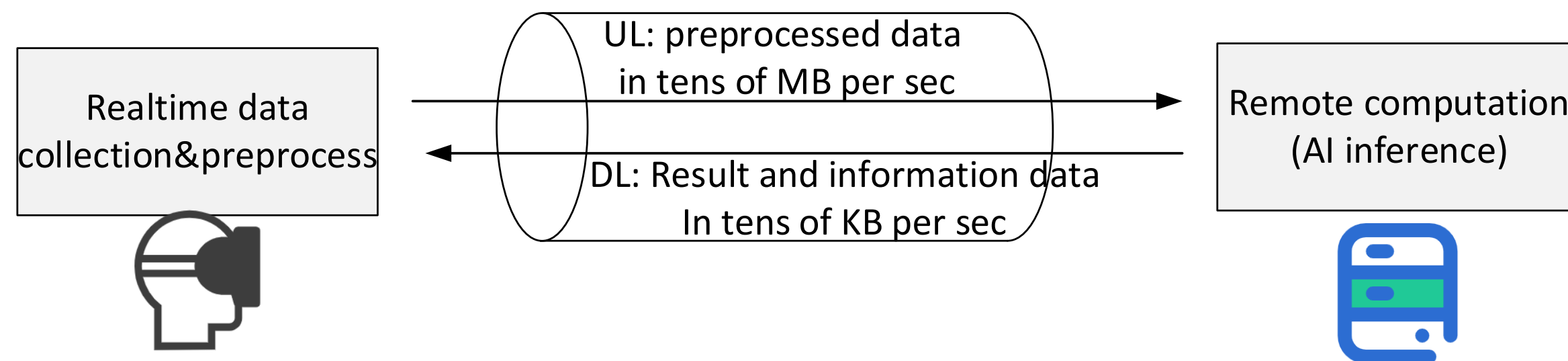
Leverage Side link for AIML operation including enhancement to Discovery, QoS, policy and service continuity

Information exposure to application in UE - scenario

- **QoS information exposure directly to UE has the following great benefits**
 - High accuracy and Fast notification to UE helps to adjust the application layer behavior in time.
 - UE need not hold a connection to application server especially saving signaling for idle mode
 - AF' s work for adaptation to 5GC can be simplified or avoided

Scenario-1: AR/VR glasses based motion-to-photon service requiring real time high reliability data transmission & response and low latency.

Scenario-2: Auto driving for in-time information (e.g. HD map) downloading and real time interaction for remote analytics



Information exposure to application in UE – Gap analysis

- **Direct exposure to UE**

Common connection for in-time message delivery especially when UE is in IDLE mode

- **Efficient Information adaption**

See AIML-Translation slide

- **Real time pre-alerting with High information accuracy**

Within allowed anticipated notification time, the network provides pre-alerting (such as QoS change) information with high accuracy. E.g. happens in 10 seconds.

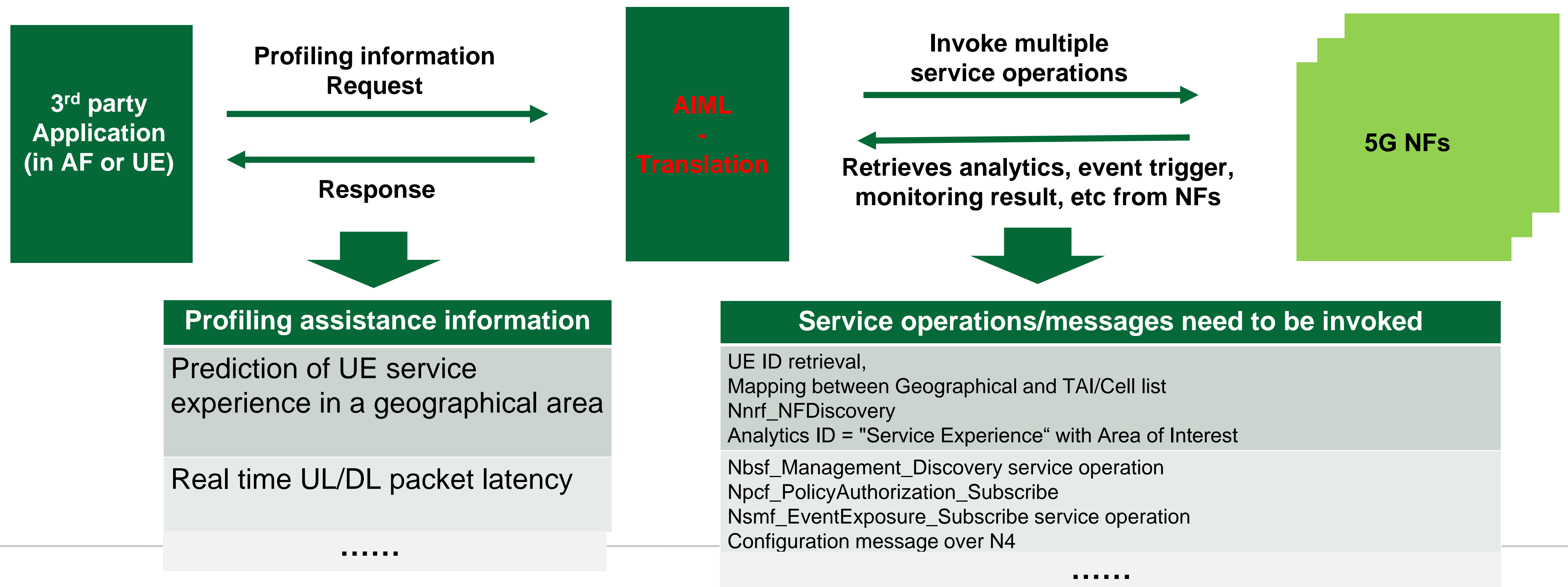
AIML-Translation between application request and 5GC APIs

Scenario

- More and more newly defined 5GC API (Service operations) makes the adaptation work complicated for a 3rd party application (both UE and AF)

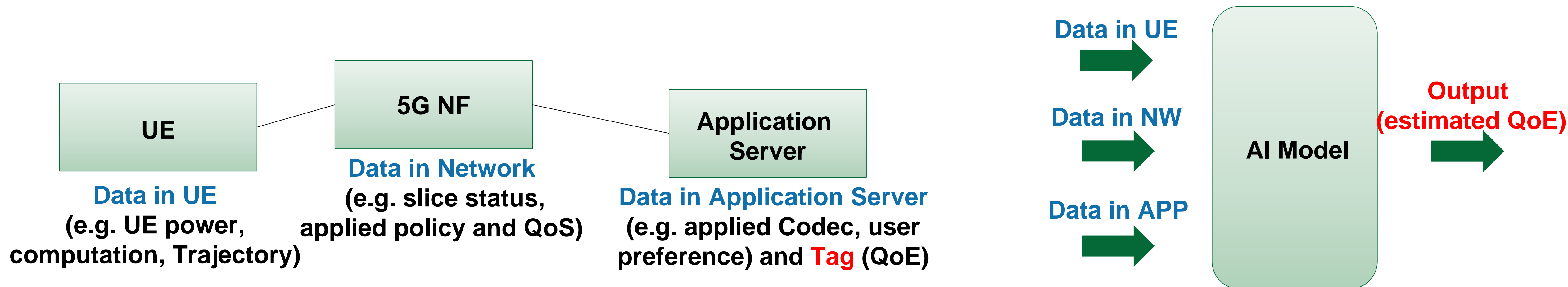
Gap analysis

- Need a mechanism to support a mapping and coordination between application request (Profiling information) and 5GC service operation. It will be much meaningful for 3rd party to adapt to acquire 5GC's information



5GS assisted Cross domain AI training and inference (Vertical FL) - Scenario

- Application layer User experience is affected by the factors in different domains, mainly related to network information (e.g. slice status, applied Policy), UE information (e.g. UE power, Trajectory) and Application layer information (e.g. applied codec, QoE metric).
- Due to some security or commercial reason, the information cannot be shared and the application layer performs AI analytics only based on application layer information
- Given usually the final target of wireless communication is to have a good QoE and the QoE is measured in application server. It may use the Vertical Federated Learning



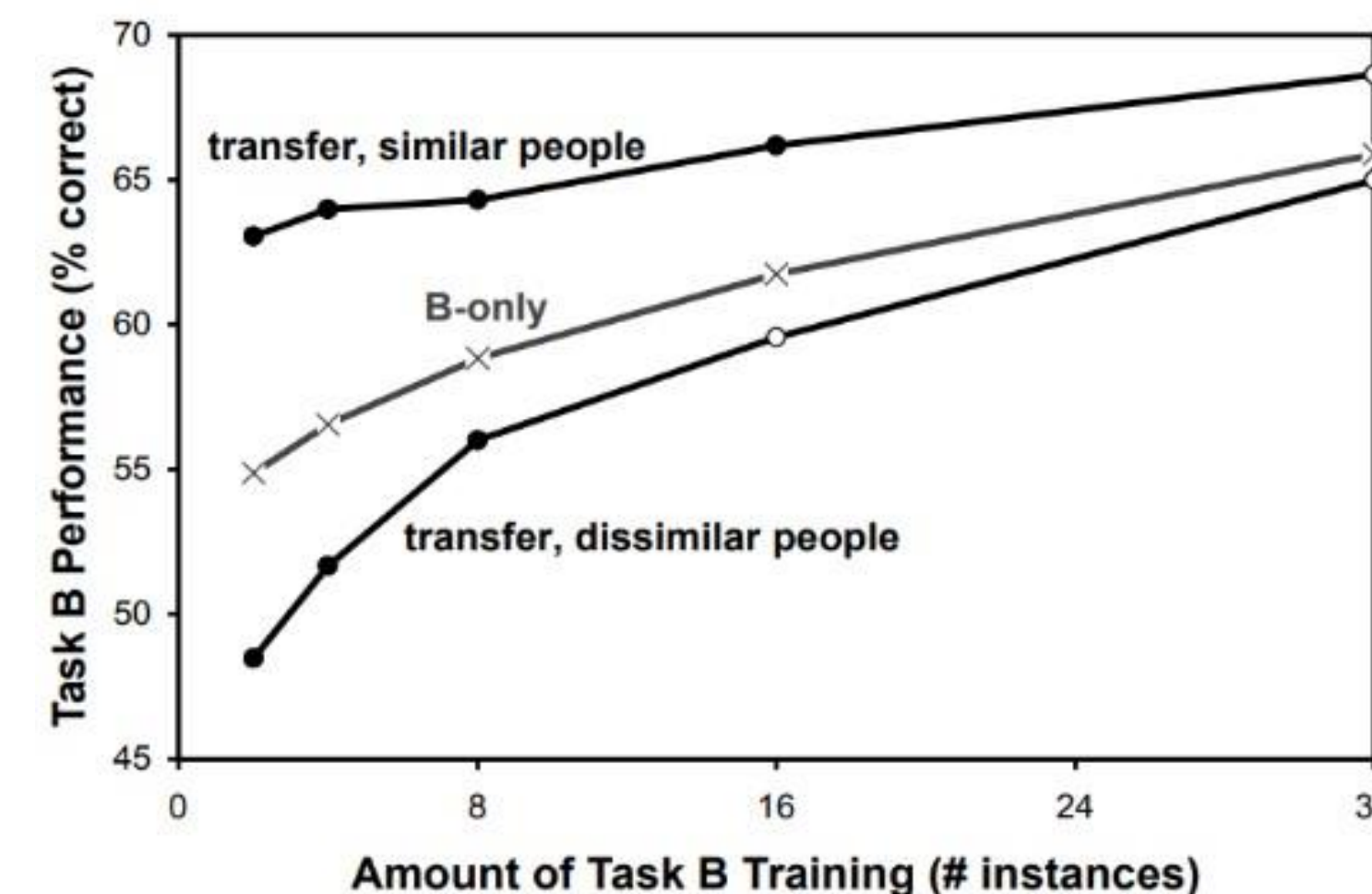
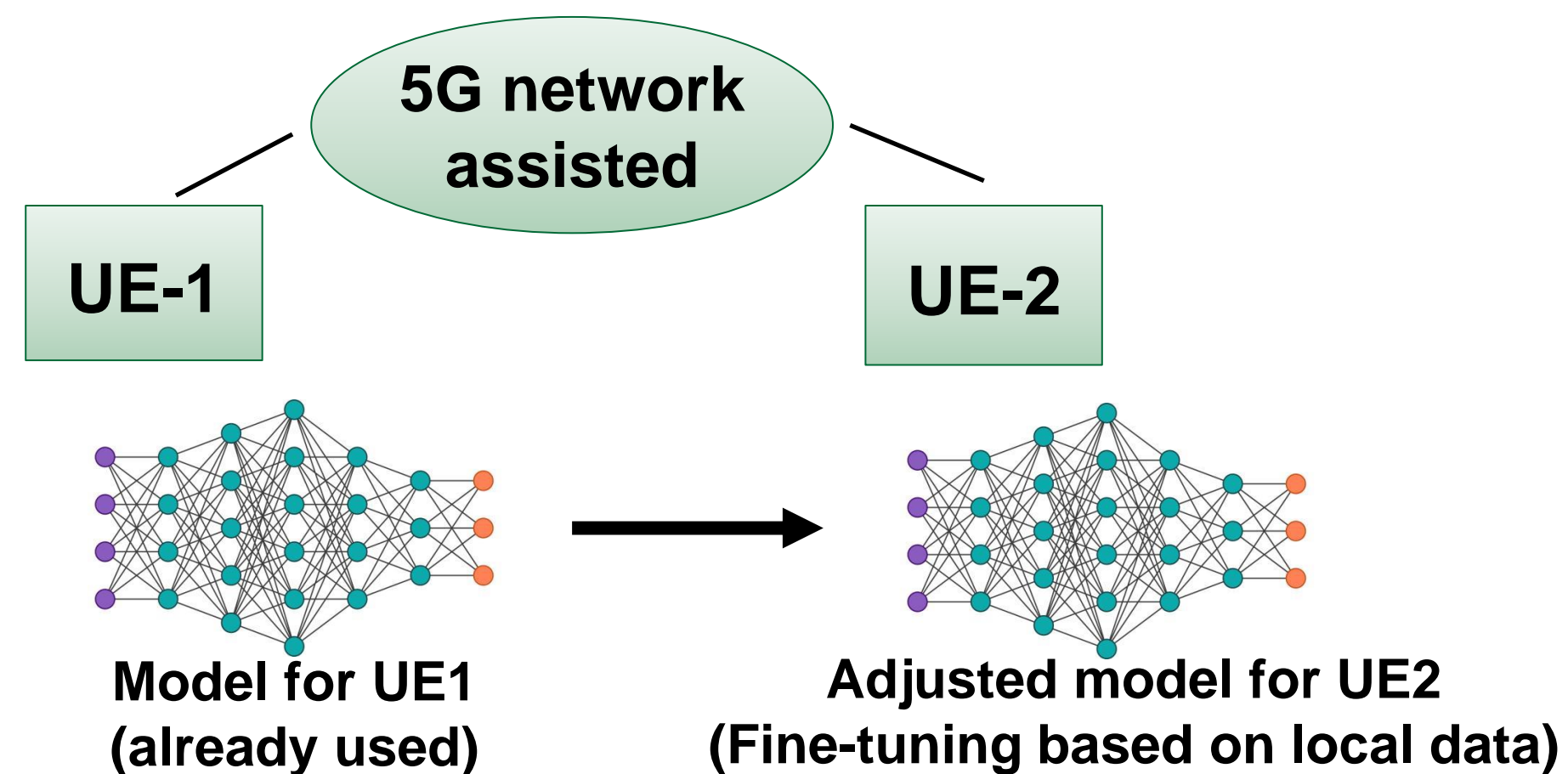
5GS assisted Cross-domain AI training and inference (Vertical FL) – Gap analysis

- The 5G NF and UE performs AI training and inference being requested by Application server. Or the network or UE as one party to organize the AI training and inference
- Data and information management for the vertical FL and inference. Such as doing Register and Discovery by AS, UE and 5G NF for a reasonable model training/inference
- The QoS and policy required for the Vertical FL and inference

5GS assisted Transfer Learning

■ Scenario

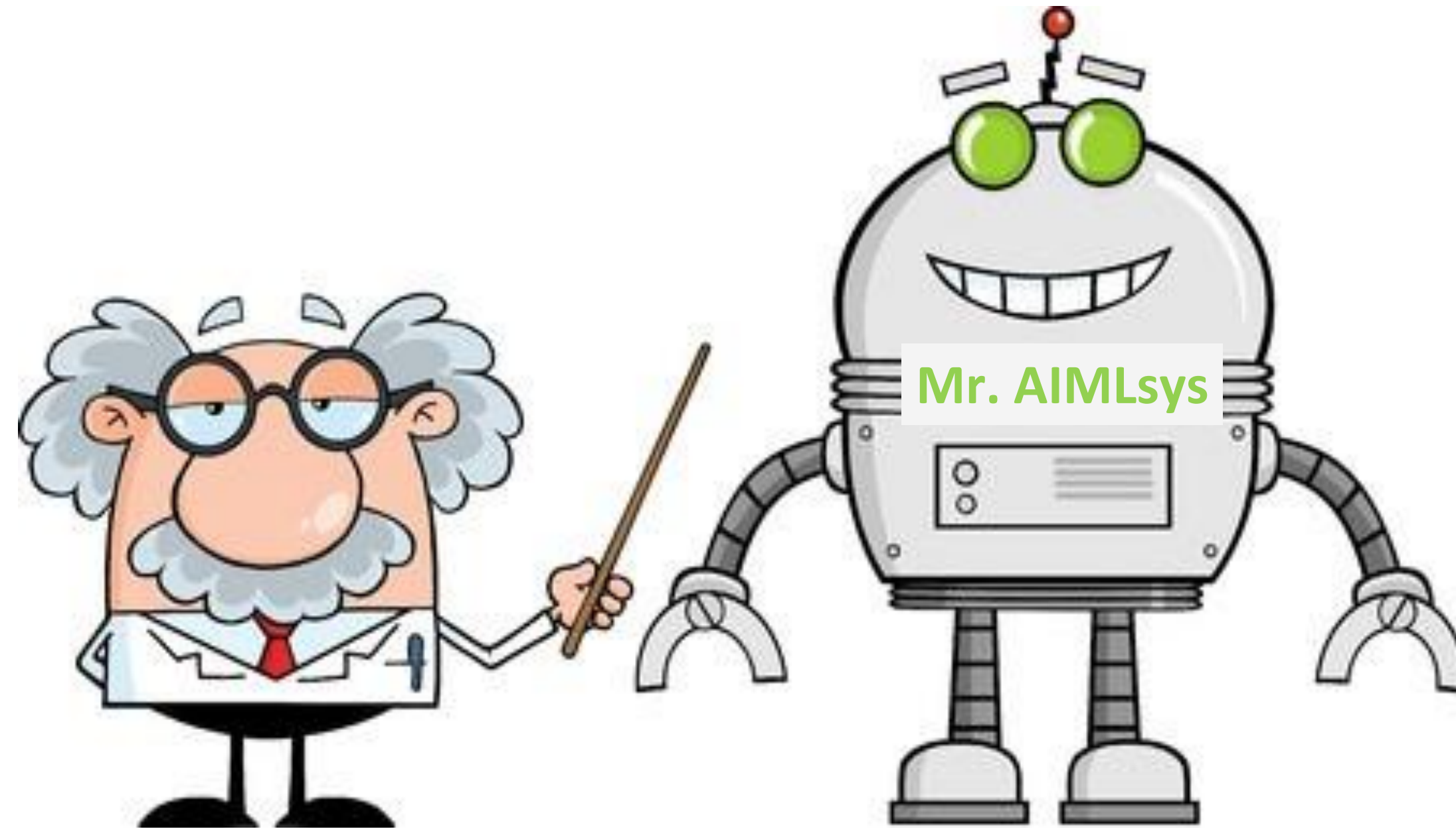
- Usually AI model is acquired from a pretrained knowledge instead of starting from scratch. Transfer learning is a key technique for efficient model training.
- Fine-tuning, as the widely used transfer learning technique, helps to train model for UEs sharing similar user characteristics/factors such as location, QoS, trajectory, applied policy, etc.



(Reference: To Transfer or Not To Transfer, Massachusetts Institute of Technology, MIT, Michael T. Rosenstein, et al.)

■ Gap analysis:

- In R18, it is defined member selection assistance functionality for federated learning. This functionality can be extended to support Transfer Learning from 3rd party
- Other policy and QoS control for the traffic for Transfer Learning needs to be studied



*The best idea is to make the world
a better place!*