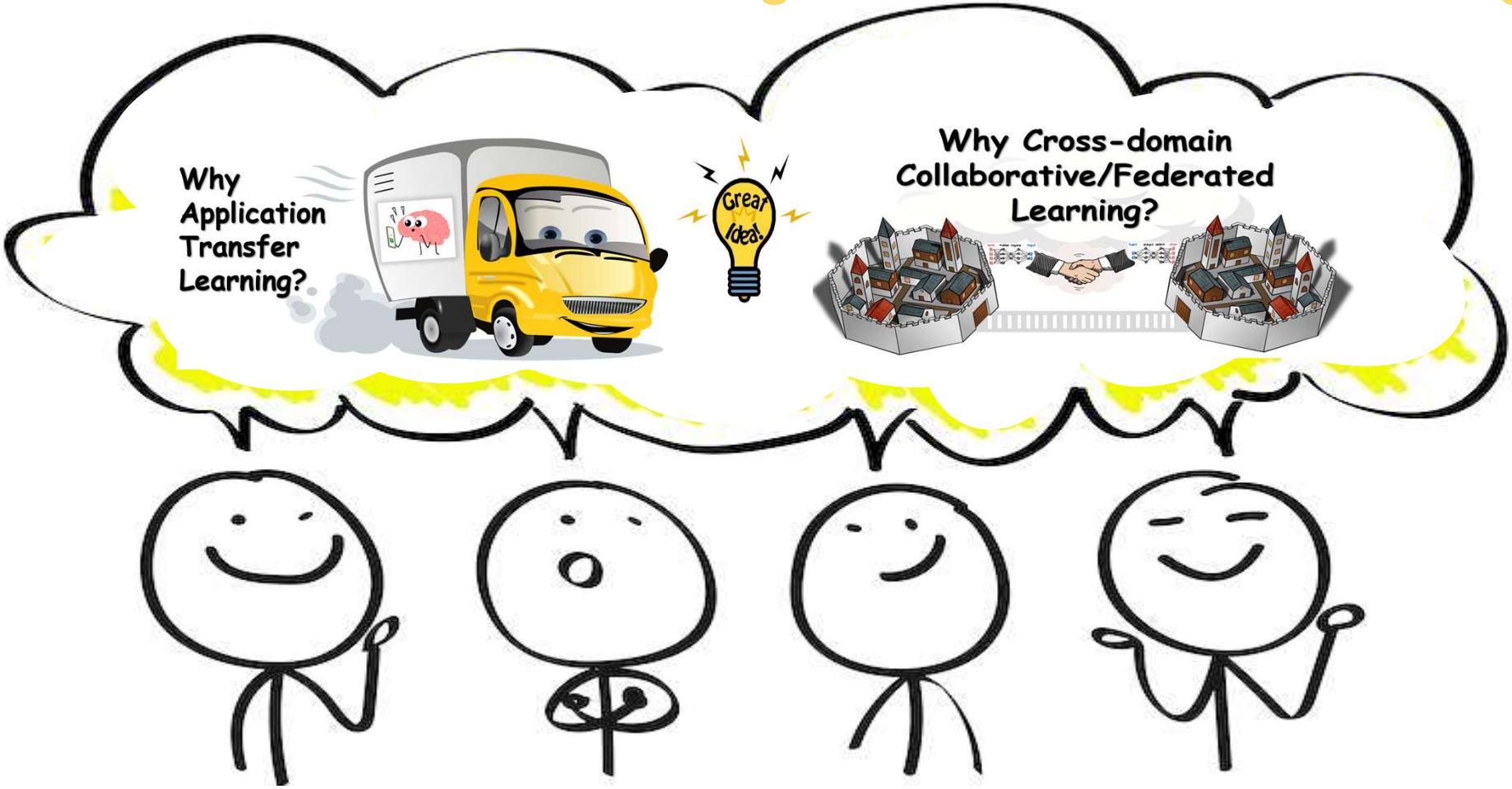
SA WG2 Meeting #S2-158
21 - 25 August, 2023, Goteborg, Sweden

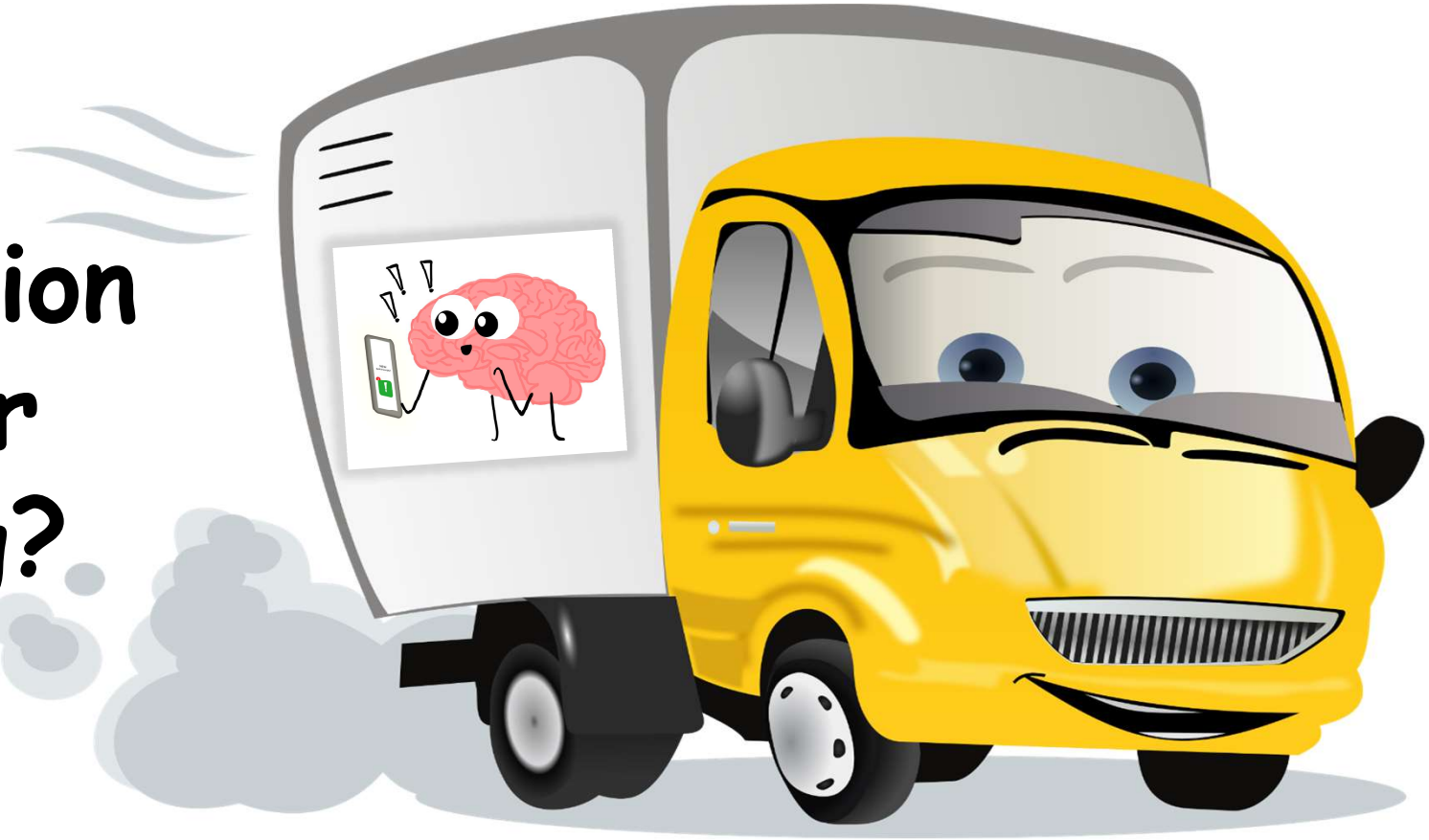
S2-2309117

Motivations of Transfer Learning and Collaborative AI Learning



oppo

Why Application Transfer Learning?



oppo

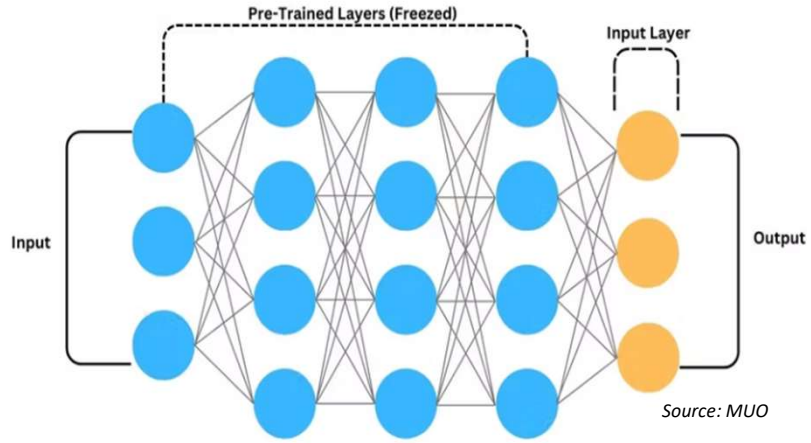
What is Transfer Learning?

Transfer learning is the idea of taking a **pre-trained model** and training it on a custom dataset to work on tasks it wasn't necessarily trained to tackle. As long as the model has developed neural networks applicable to the model you're trying to train, you can use just about any model for transfer learning.

Transfer learning provides several **benefits** over training an AI model from scratch.

- √ **Reduced Training Time:** When training a model from scratch, a large part of the training process is spent on general foundational knowledge. Through transfer learning, your model automatically inherits all of this foundational knowledge, thus significantly reducing training time.
- √ **Lessen System Resource Requirement:** Since all the foundational knowledge is already there, all you need to do is to further train the model for the specifics of your application. This often only requires a relatively small data set (i.e. lessen resources for data collection) that can be processed with much less computing power, system capacity and airlink resources to ensure the service continuity support for data transfer.
- √ **Improved System Performance:** Unless you spend millions of dollars on building your model from scratch, you cannot expect a model as good or reliable as a large language model (LLM) from a giant tech company. By using transfer learning, you can leverage the powerful capabilities of these pre-trained LLMs, similar to the concept of GPT, to enhance your model's performance.
- √ **Energy Efficiency:** With three main benefits described above, it also translated into more energy saving on the device, the network and the application server.
- √ **Lower OpEx & CapEx:** With the four benefits described above, it also translated into lower operation cost for MNO. In fact, due to lesser system capacity and network resources, it may also translate into lower CapEx for the capital investment.

High-level Concept of Transfer Learning



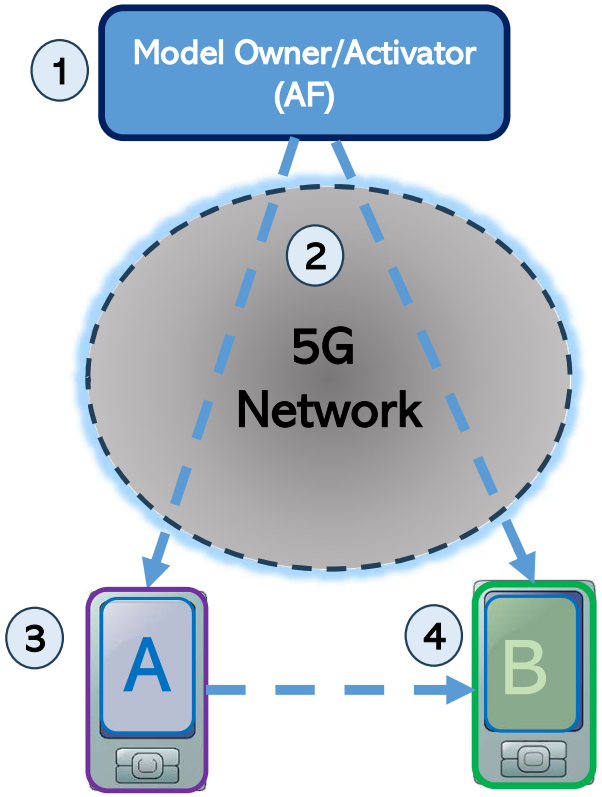
Source: MUO

Training an AI model from scratch is possible, but you need greater resources to do so and this implies:

- X ***More** training time*
- X ***More** system resource requirements*
- X ***Lower** system performance*
- X ***Energy** Inefficiency*
- X ***Higher** OpEx and CapEx*

oppo

EXAMPLE for How Transfer Learning Works & How it could impact 5GS?

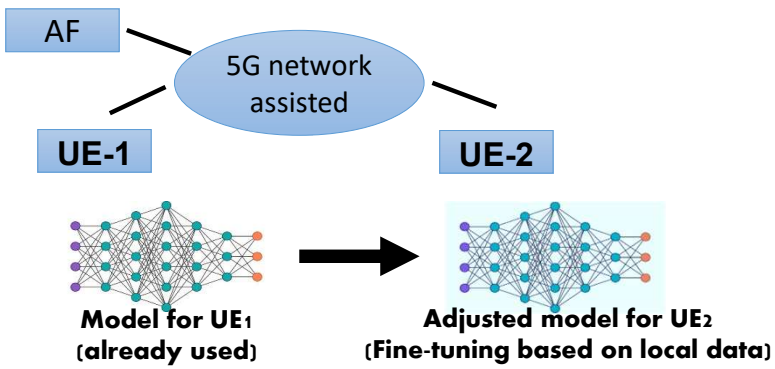


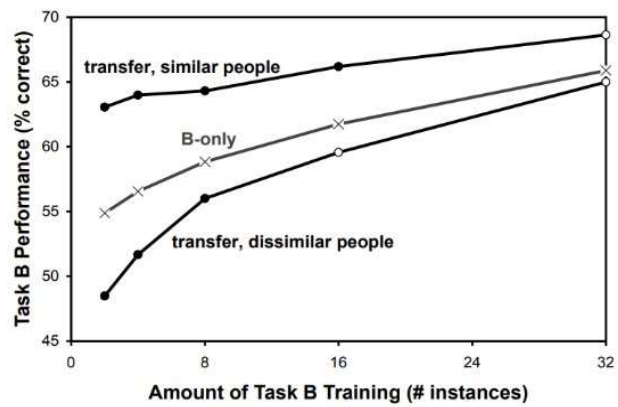
Transfer Learning Operation	AI/ML Operation Descriptions	Entit(ies) involved in the Operation	3GPP Impact
1 Selecting pre-trained Model	AF selects pre-trained Model from a source task corresponding to a set of features (i.e. A, B, C, D) which have a lot of similarities of the new target training model.	AF	None
2 Selecting Source/Target UEs	AF leverages existing Restricted Discovery Feature from eProSe and extends existing Member UE Selection features from AIMLsys to select the source and target UEs for Transfer Learning.	AF, 5GC, UE	Extensions to AIMLsys for Member UE Selection to select the proper source and target UEs.
3 Initiating Transfer Learning	Once the AF receives the confirmation that source and target UEs have successfully connected with each other, it will trigger the Model Transfer **at the application layer**.	AF, UE	None
4 Model Fine-tuning	Freezing the initial layers of the pre-trained model (closest to the input) to keep their weights during fine-tuning. Removing the final layer of the model to enable it to be trained for the new application. Applying new dataset to the learning algorithm, which is then used to train the new model and the specifics of its application	UE	None

oppo

Why 5GS Assistance for Application Transfer Learning for R19?

- √ SA1 TS 22.261 has well defined system requirements for Transfer Learning for R19
- √ Transfer learning requires less training time and less system resources, while it offers better system performance, more energy efficient and lower OpEx and CapEx
- √ Low hanging fruit (i.e. minimum TU) to achieve maximum benefits to enable efficient and effective 5GS assistance for Application AI/ML support

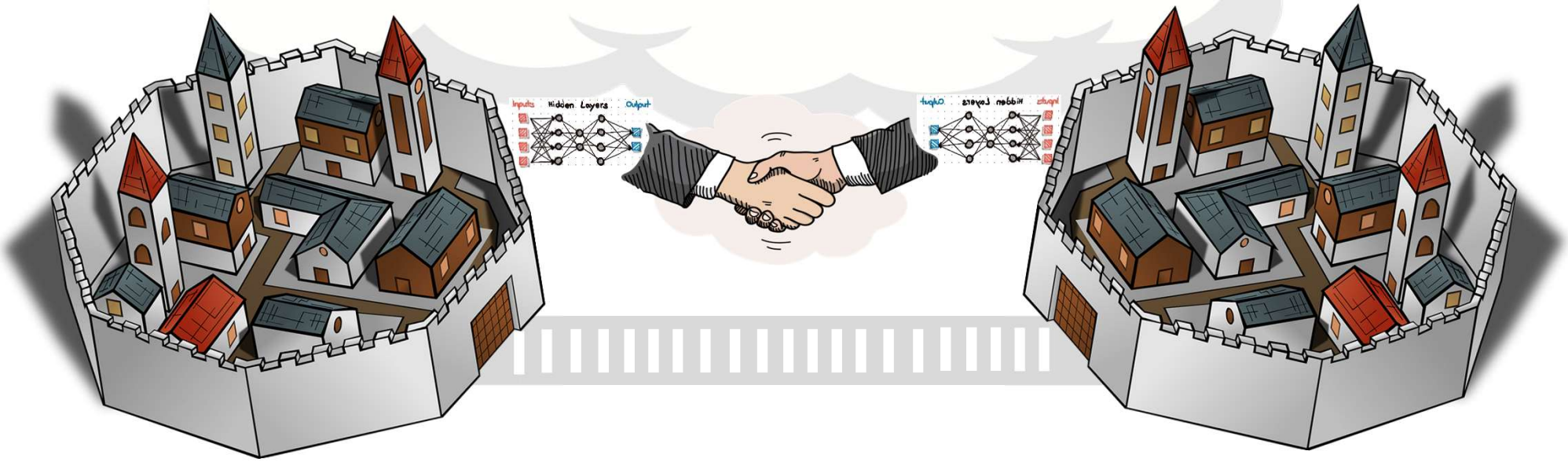




(Reference: To Transfer or Not To Transfer, Massachusetts Institute of Technology, MIT, Michael T. Rosenstein, et al.)

oppo

Why Cross-domain Collaborative AI Learning?



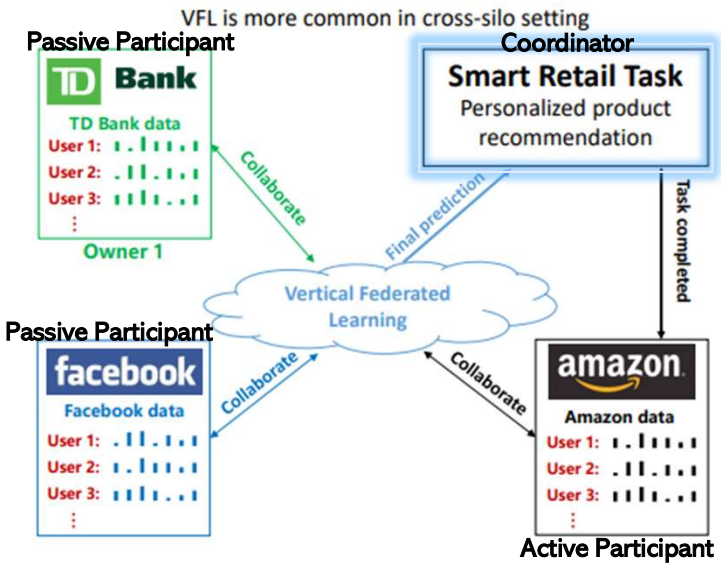
oppo

What is Cross-domain Collaborative AI Learning?

Cross-domain Collaborative AI Learning is to apply the Vertical Federated Learning (VFL) concept which is a distributed machine learning technique to leverage distributed local data corresponding to the same subject/target (e.g. UE) but different feature spaces (e.g. mobility, QoS, availability etc.) among the training participants (i.e. domain of the local dataset, e.g. 5GC, OAM etc.). In many more practical scenarios, VFL is more relevant than Horizontal FL (HFL) because different domains (e.g., 5GC, OAM) hold different features (e.g., QoS, Mobility Pattern etc.) for the same set of users.

Examples of cross-domain collaborative AI Learning

No local data exposed out of each domain



- **Smart retail task:** as shown in the figure, a **retailor** (e.g., Amazon) wants to improve the performance of personalized product recommendation by using the financial data from a **bank** (e.g., TD Bank) and the social relation data from a **social network company** (e.g., Facebook).
- **Credit score evaluation task:** the bank wants to improve the evaluation accuracy of user credit scores by using the purchasing data from the retailor and the social relation data from the social network company.
- **Content recommendation task:** the social network company wants to push user interested contents more accurately by analyzing the financial data from the bank and the purchasing data from the retailor.

Source: SIGKDD 2021

In VFL, there are typically three types of participants, namely active participants, passive participants, and a coordinator.

Active participant: An active participant is an entity in VFL that wants to build models and provide data with a sample set {si , · · · , sj} and their labels {yi · · · , yj} (i.e. target outputs).

Passive participant: A passive participant is an invited entity in VFL that provides data with the sample set {si , · · · , sj} but with no labels, and its feature set is different than that of the active participant. There can be multiple passive participants in VFL.

Coordinator: A coordinator coordinates the training process in VFL, and it communicates with the participants in an encrypted channel. It has no access to the raw data. In real-world scenarios, the role of the coordinator is usually performed by the active participant or a trusted third party

oppo

How 5G Cross-domain AI Learning Works? (Use Case: Application QoE Prediction)

QoE has been defined by ITU-T SG12 (2007) as *"the overall acceptability of an application or service, as perceived subjectively by the end user"*, and it is closely related to the specific service delivery. It includes the total end-to-end system effects, and it may be influenced by user expectations and degree of delight or annoyance during the fruition of a service

Problem Statement:

- It is difficult to express QoE in an objective and mathematical way, because it can depend both from the context in which the user operates and both from the quality negotiated with the network operator for the specific service.
- Historically, the quality has been made by means of network monitoring in order to extract and calculate network performance parameters like packet delay, jitter, packet loss, without referring to the particular service. Such an approach presented a critical limit.
- How the network operator reacts in real time in order to solve the cause of the worsen QoS indicated by the performance indicators, because these was available for the analysis only at the end of the service provisioning. In addition, this approach didn't take into account of the user experience.
- NWDAF relies some inputs data from the AF on the QoE Metrics to predict the User Experience; however, whether AF is going to share such info with MNO and how AF obtains the future QoE info to support NWDAF is unclear.

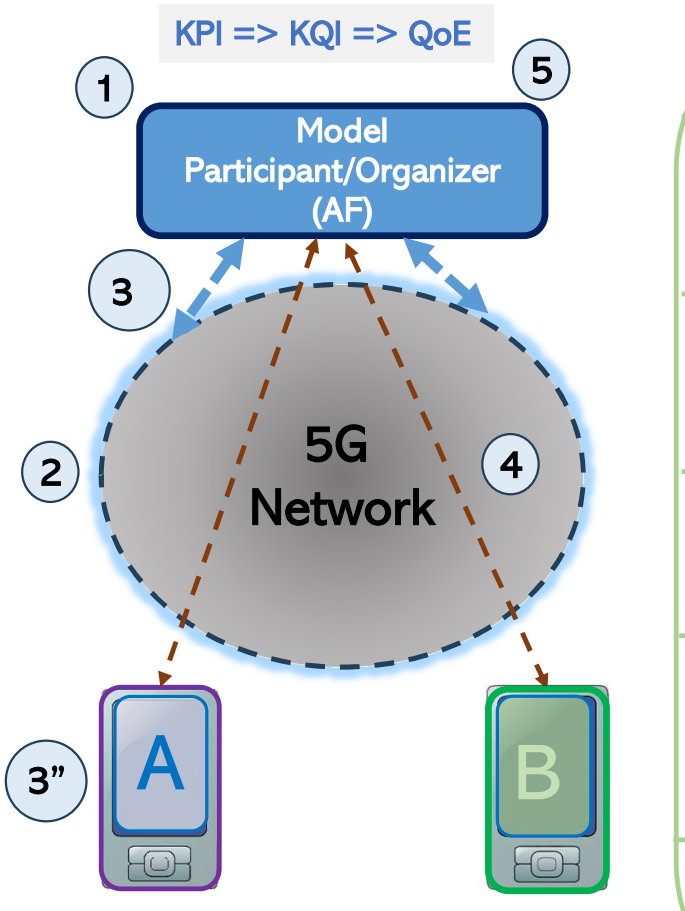
EXAMPLE of Illustration on how to address this Application QoE Prediction Issue:

It is proposed to establish a relation between the QoS metrics and the user expectations (referred as QoE performance metrics). QoS metrics are considered as the ability of the network to provide a service at a guaranteed performance level. They are assured by the differentiated management of the various traffic classes, and it is a technical, objective and network oriented entity, mathematically expressed by network parameters, obtained by monitoring operations on the access network, core network and end user device according to the KPIs which are specified for them. QoE performance metrics have been understood as the reflection of the user expectation about the delivery of service and is referred as KQI. By specifying how to map the KPI to KQI, it is then feasible to identify the QoS/QoE relationship, and therefore to derive the QoE.

NEXT: Examining the use case of deriving the Application QoE Prediction for 5G Voice and Video Streaming services via the support of Cross-domain VFL operation to leverage the KQIs of the required features for 5G Voice and Video Streaming services.

oppo

EXAMPLE for How Cross-domain Collaborative AI Learning Works & How it could impact 5GS?



$QoE = f(KPI_1(QoS_1), KPI_2(QoS_2), \ldots, KPIn(QoSn))$

Cross-domain FL Operation	AI/ML Operation Descriptions	Entit(ies) involved in the Operation	3GPP Impact
1 Cross-domain FL preparation	AF determines the QoE (i.e. Label) for the target application and the required features in each domain (e.g. UE, 5GC) and the corresponding KPIs for the features	AF	None
2 Selecting FL Member UEs	AF may directly or leverage the Member UE selection criteria to 5GC to select one or more UEs to perform the cross-domain FL.	AF, 5GC,	Re-use existing Member UE selection
3 Initiating Cross-domain FL operation	AF triggers the cross-domain FL operations towards the 5GC and UE by sending the local training model, the features and the KPIs corresponding to the specific target domain.	AF, 5GC, UE	Extend 5GC Service Provisioning Note that, AF communication with UE should be at the application layer.
4 5GC local FL operation	5GC performs the feature mapping to select the proper target NF(s) to collect the required dataset corresponding to the features and the Member UEs. It then triggers the local training/inference operation by referring to the KPIs.	5GC	NF selection and performing local training (e.g. at NWDAF) by referring to the specified KPI. UE's local training is at application layer.
5 Model Update/ Inference report	AF collects the intermediate data or inference results to perform the training/inference aggregation to derive the QoE. AF may trigger another round of FL operation.	AF	None

oppo

EXAMPLE of KPI/KQI Video Streaming Service Relationship

SERVICE KPI (KQIs)	Network KPI (3GPP TS 32.450, TS 32.455)								
	Accessibility	Retainability	Integrity				Cell Availability	Mobility	Resource Utilization
			Throughput	Delay	Jitter	PLR			
CONNECT TIME (Session Setup)	✓						✓		
AVAILABILITY (Session Active)		✓					✓		
VIDEO/AUDIO QUALITY (Session Active)				✓	✓	✓		✓	✓
RESPONSE TIME (Session Active)		✓		✓	✓				✓

oppo

EXAMPLE of KPI/KQI Voice Service Relationship

SERVICE KPI (KQIs)	Network KPI (3GPP TS 32.450, TS 32.455)									
			Integrity							
	Accessibility	Retainability	Throughput	Latency	Jitter	PLR	Cell Availability	Mobility	Resource Utilization	SIR
SERVICE AVAILABILITY	✓	✓					✓	✓		
SPEECH QUALITY (MOS)				✓	✓	✓				✓
END TO END DELAY	✓			✓	✓					

oppo

EXAMPLE for WHY Cross-domain Collaborative AI Learning for Application Energy Efficiency Prediction?

Energy Efficiency should not ONLY benefit 5G system. ASP could also benefit from energy efficiency for its own application service and for its 5G devices in order to ensure the UE battery life for running the application without sacrificing the QoE. Therefore, it is advantageous for the AF to initiate the parallel AI training/inference operations of the cross-domain collaborative FL between the AF and the 5GS to predict both the energy efficiency and QoE for its services w.r.t. target UE(s). Once the AF verifies both aspects, it could also assist the 5GS to make intelligent policy decisions such as the following:

Reduced Distance to Base Stations

The power required to communicate between clients and 5G base stations increases the further the signal has to be transmitted. Since small cell base stations are deployed in close proximity to client devices, it significantly reduces power consumption by both the base stations and the 5G client devices.

Increased Network Density

Small cells with massive MIMO antennas can serve many more devices at the same time. Each device is multiplexed over the same space and frequency. This spatial multiplexing uses the same channel to serve multiple devices. The energy consumption is also shared among multiple users or devices.

EXAMPLE for How Cross-domain Collaborative AI Learning for Energy Efficiency Predictions Use Case?

- **AF leverages the Cross-domain Collaborative FL to collect the following parameters to determine the Energy Efficiency (per UE, per service or per slice) corresponding to its service application.**

Energy consumption KPI from OAM.

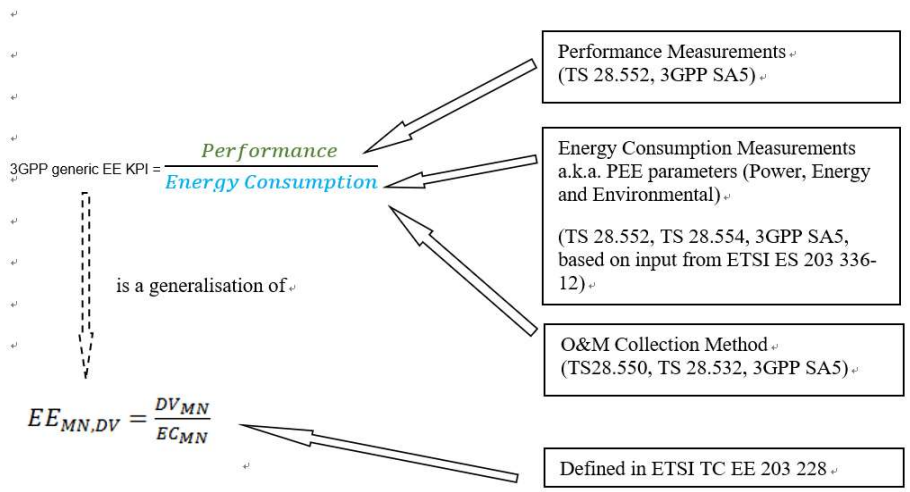


Figure A.2-1: KPI derivation and sources

Power, Energy and Environmental (PEE) measurements:
watts (W), kilowatt-hours (kWh) degrees Celsius (°C). volts (V). amperes (A) Humidity

parameters	Domain
NF load	5GC
Data usage for the specified application	App
QoS performance	5GC
The adoption of alternative sustainable sources of energy	5GC
Whether to agree to chose the green communication service	UE
RAN Coverage quality.	RAN
NG-RAN deployment scenarios（Dual connection, CU-DU）	RAN

oppo

Why Cross-domain Collaborative AI Learning for R19?

√ Data Security and Privacy Protection: Keep training data locally at the training domain (e.g. UE, RAN, 5GC, AF etc.) and lessen the concern to comply to GDPR-like regulations.

√ Data diversity: Heterogeneous data because of using data from different domain specific datasets.

√ More applicable and more practical for Mobile System compared to Horizontal FL (HFL) because different domains (e.g., 5GC, OAM) hold different features (e.g., QoS, Mobility Pattern etc.) for the same set of users.

√ Decentralized ML with smaller datasets can further reduce the network bandwidth and energy consumption by sending only the features of interest rather than the stream of the raw data.

√ Expanding the MNO service offering to ASP which requires 5GC to provide more specific assistance (i.e. leveraging 5GC to perform local training over its local dataset) to assist specific ASP service application which requires sensitive local data from other domains.

- E.g. Application Service provider requests MNO to assist its service domain for Energy Efficiency training, ASP can specify the required features and KPIs to 5GC to assist the ASP to perform the energy efficiency training operation. .

oppo