

AI in Media Services

3GPP

About the project

The reports the evaluation framework is built around

3GPP TR 26.847

Technical Report, V19.0.0

Evaluation of AI and ML in 5G media services: the evaluation testbed architectures and anchors, the metrics and the evaluation scenarios.

3GPP TR 26.927

Technical Report, V19.0.0

Study on AI and ML in 5G media services: the media use cases and the architectures for AI/ML model distribution and split operation.

3GPP TR 22.874

Technical Report, V18.2.0

Study on traffic characteristics and performance requirements for AI/ML model transfer in 5GS.

FIND OUT MORE

Technical Analysis

Specifications

Testbeds

The evaluation framework

What TR 26.847 defines for the evaluation of AI/ML in media services

Testbed architectures clause 4, common to every scenario unless stated

Anchors: inference on the device, or in the network

Split inference intermediate data testbed

Model data testbed

Metrics clause 4.4, with the AI/ML task metrics

AI/ML task metrics

Video quality

Complexity

Bitrate

Split model size

Intermediate data size or bitrate

Compressed model size

Compressed intermediate data ratio

Latency

Scenario template annex A, one entry per evaluation scenario

Motivation and use case relevance

Description

Anchor model

Testbed architecture and anchors

Test configuration

Metrics and requirements

Test dataset and scripts

Test conditions

Interoperability

The functional side of AI/ML in media services is documented in TR 26.927; TR 26.847 is its evaluation companion.

Scenarios

The four evaluation scenarios of TR 26.847 and the scenarios of the repository

In TR 26.847 the four evaluation scenarios of clause 5

Compressed AI/ML model data for speech recognition

Split inferencing for object detection and labelling

Bit-incremental transmission and deployment of AI/ML models

Real-time translation over IMS

Scripts in the repository read from the README and the scripts folders

Compressed model transmission for speech recognition

Split inferencing for object detection

IMS real-time translation

Planned described in the README, with no scripts yet

Split inferencing for human pose estimation

Split inferencing for hand gesture recognition

Video super resolution in video call

Federated learning

The repository is an evaluation tool under development, not a service or a network function.

What the scripts do

The three scenarios that have scripts today

COMPRESSED MODEL TRANSFER

Speech recognition models

A pre-trained model (wav2vec 2.0 or HuBERT) is compressed before it is sent to a device. A codec pipeline encodes and decodes it and reports its size and its recognition performance before and after.

SPLIT INFERENCE

Object detection

A pre-trained model detects objects in a video sequence and is split in two parts, one on the device and one in the network. Scripts convert and split a RetinaNet model and run it whole or split.

TRANSLATION

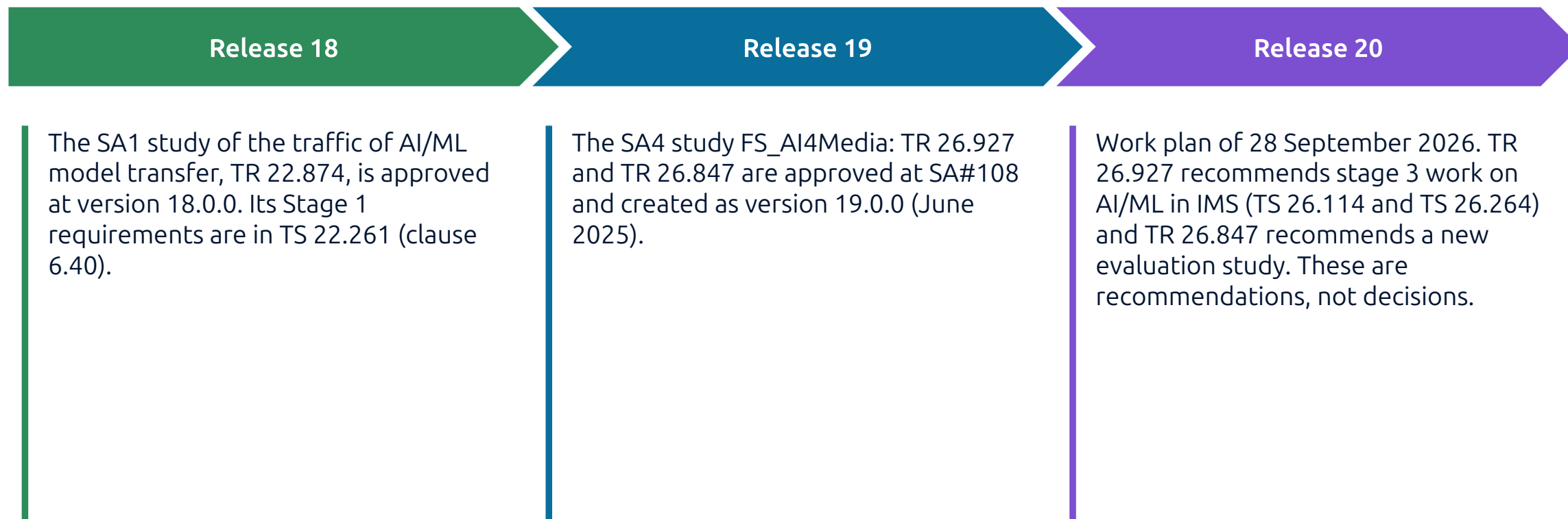
Translation models

Scripts train a Transformer translation model on Tatoeba sentence pairs and translate sentences, with split translation and a WebRTC translator. Four trained models are stored with Git LFS.

A Dockerfile builds the evaluation environment (Ubuntu 22.04, with Nvidia GPU support). The datasets are hosted outside the repository.

How it grew

What each 3GPP release added to AI and ML in media services



From the Standards pages of the website, and clause 7 of TR 26.927 and clause 6 of TR 26.847.

IMPLEMENT

Repositories

The repository of the AI/ML evaluation framework

REPOSITORY	WHAT IT IS	COVERS	LATEST RELEASE
rt-ai-ml-evaluation-framework	the implementation of the AI/ML evaluation framework defined in 3GPP SA4 TR 26.847 (Release 19)	Compressed model transfer Split inferencing Translation scenario	● On request Early Access

FIND OUT MORE

Technical Analysis

Specifications

Testbeds

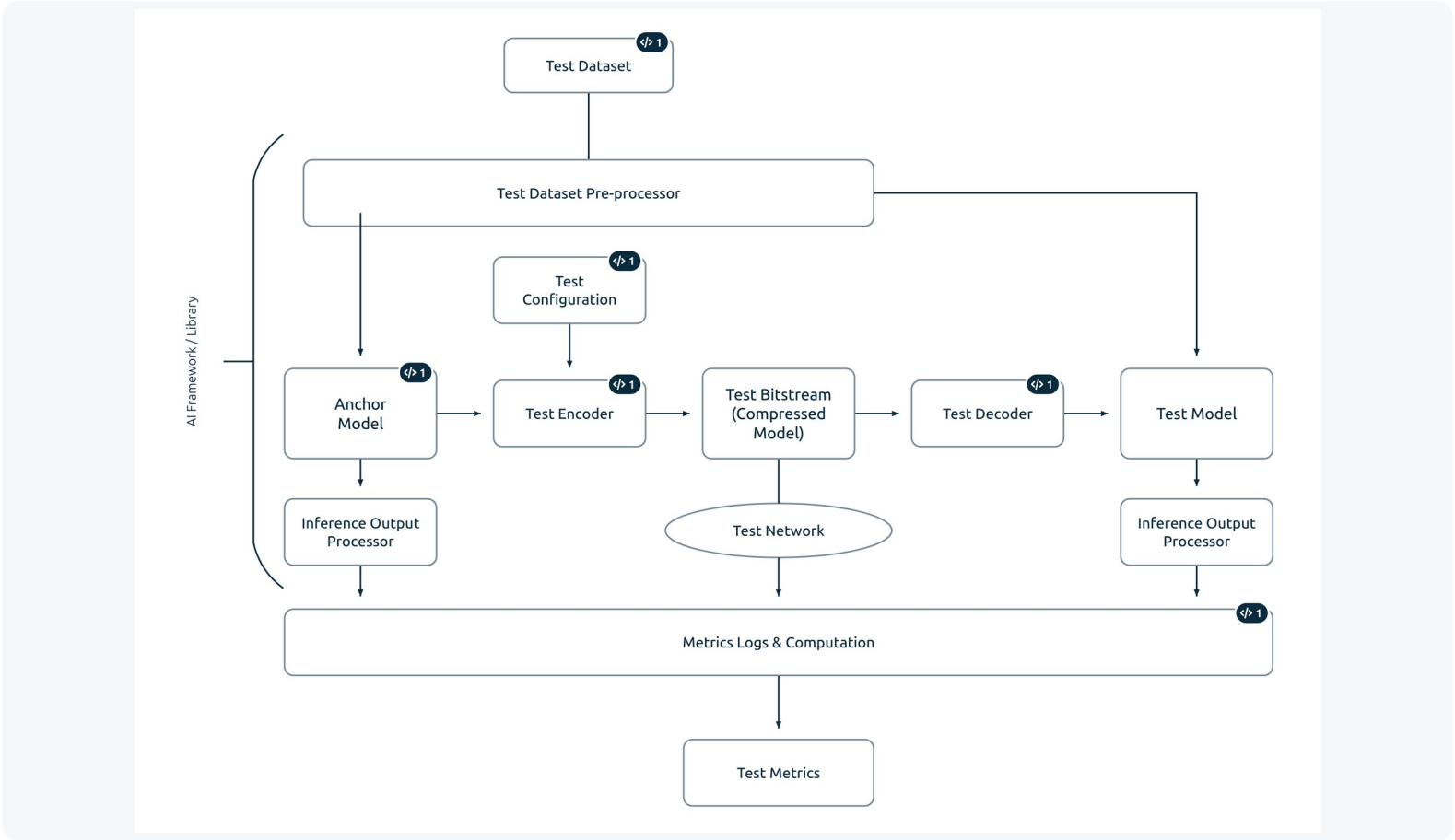
● released

● no release yet

IMPLEMENT

Repositories in the architecture

The Reference Tools on the specification figure, as the figure generator draws them



PROVIDED BY THE REFERENCE TOOLS

- 1 rt-ai-ml-evaluation-framework

Boxes without a badge are not mapped to a repository.

Latest updates and roadmap

Status on 23 October 2026

NO SCRIPTS YET

- **Human pose estimation**
The README describes the split inferencing of PoseNet at different layers; the repository has no scripts for it yet.
- **Video super resolution**
The README gives the motivation: video calls are typically 480p today and XR calls need more. The repository has no scripts for it yet.
- **Hand gestures and federated learning**
The README has the headings and no text. Both are planned.

IN DEVELOPMENT

- **Bit incremental transfer**
A pull request is open to merge the bit incremental transfer feature into the main branch.



See it. Try it. Contribute.

Raise an issue, send a pull request or join the monthly public call.

[Become a contributor →](#)



GitHub



5g-mag.com

CONTRIBUTORS

