TS.53 AI Mobile Device Specification Test book

Version 1.0

26 July 2023

Security Classification: Non-confidential

Copyright Notice

Disclaimer

Compliance Notice

The information contain herein is in full compliance with the GSM Association’s antitrust compliance policy.

This Permanent Reference Document is classified by GSMA as an Industry Specification, as such it has been developed and is maintained by GSMA in accordance with the provisions set out in GSMA AA.35 - Procedures for Industry Specifications.

Table of Contents

1 Introduction

1.1 Overview

While more and more device manufacturers declare their products support AI, unfortunately there are obvious difference in the functionality and performance, which greatly impact the user's experience and perception of AI. In order to guide the industry and align the performance of AI mobile device, GSMA has published TS.47 to set up the standard for AI mobile device. The purpose of this document is to define test cases to verify the compliance of a device to the requirements defined in GSMA PRD TS.47.

1.2 Scope

This document defines the test cases to verify whether a device comply to GSMA PRD TS.47, including test set-up, test method, test procedure and expected results.
The devices covered by this document are mobile devices and tablets.

1.3 Definition

Term	Description
AI Mobile Device	Refer to the definition of AI Mobile Device in TS.47 AI Mobile Device Specification [1].
Power Meter	Equipment that used for power measurement and can fulfil the following functions: 1. Provide power for DUT. 2. Display the current value in diagram.
VGG16_notop	VGG16 without last three fully connected layers, in float32 format. [7]
Audit	"The audit privilege (including source code access) should be given/conducted by auditor(s) who may need to sign an Non- Disclosure Agreement with the OEM. However the OEM cannot refuse a legitimate request to audit the source code."

1.4 Abbreviations

Term	Description
AI	Artificial Intelligence
OEM	Original Equipment Manufacturer
DUT	Device Under Test
SDO	Standard Developing Organisations
TOPS	Tera Operations Per Second
TOPS/w	Tera Operations Per Second / Per Watt

1.5 References

Requirements shall be based on the exact versions as indicated below. However if the manufacturers use a later release and/or version this should be indicated. The GSMA will continually align with other SDOs as appropriate.

Ref	Doc Number	Title
[1]	GSMA PRD TS.47	AI Mobile Device Specification, Version 1.0, September 2019
[2]	ISO-IEC-19795-1	Information technology — Biometric performance testing and reporting —Part 1: Principles and framework
[3]		FIDO Biometrics Requirements https://fidoalliance.org/specs/biometric/requirements/Biometrics-Requirements -v2.2-fd-20211206.pdf
[4]		Void
[5]	RFC 2119	“Key words for use in RFCs to Indicate Requirement Levels”, S. Bradner, March 1997. Available at http://www.ietf.org/rfc/rfc2119.txt
[6]	RFC8174	Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words https://www.rfc-editor.org/info/rfc8174
[7]		https://keras.io/api/applications/vgg/
[8]	ETSI EG 202 396-1	Speech and multimedia Transmission Quality (STQ); Speech quality performance in the presence of background noise; Part 1: Background noise simulation technique and background noise database https://www.etsi.org/deliver/etsi_eg/202300_202399/20239601/01.02.04_60/eg_20239601v010204p.pdf

1.6 Modal verbs terminology

The key words “MUST”, “MUST NOT”, “REQUIRED”, “SHALL”, “SHALL NOT”, “SHOULD”, “SHOULD NOT”, “RECOMMENDED”, “MAY”, and “OPTIONAL” in this document are to be interpreted as described in RFC 2119 **Error! Reference source not found.** (RFC8174) [6] when, and only when, they appear in all capitals, as shown here.

2 Test Methodology

2.1 Testing of optional requirements

Any requirement which is optional may be subject to a conformance test if it is supported by the DUT.

A declaration by the device manufacturer based on Applicability Table (Annex A), is used to determine whether an optional requirement is supported.

2.2 Void

2.3 Repetition of tests

As a general rule, the test cases specified in the present document are highly reproducible and don't need to be repeated unless otherwise stated.

2.4 Testing of cases that can leverage the existing certification scheme

For those test cases that can leverage the existing certification scheme, the corresponding conformance test here to will directly accept the results from the certification scheme.

3 Hardware Performance

For guidance, some testing methods are provided in **Error! Reference source not found.**.

Unless stated otherwise, the testing methods defined in **Error! Reference source not found.** can be used with test cases defined in this section.

For each test case, the DUT is configured as required by the test method.

3.1 Integer implementation performance

Test purpose

To verify the DUT can meet the minimum requirements of int8 TOPS and int8 TOPS/w.

Referenced requirements

TS47_3.1_REQ_001	An AI Mobile Device SHOULD have a minimum of (1) int8 TOPS.
TS47_3.1_REQ_003	An AI Mobile Device SHOULD have a minimum of (0.5) int8 TOPS/Watt.

Preconditions

The output of converted model should closely match the output of reference model based on proposed approach in the Annex E. Scripts to pre-process the test dataset, run the test model and measure TOPS.

AI Application for TOPS testing shall have the following characteristics:

a) Source code must be available for software auditing, and
b) AI application shall support the NN baseline model used to evaluate TOPS performance, and
c) AI application shall create a test report with TOPS configuration and performance.

Test Dataset

Test dataset shall be publicly available. Example of public datasets provided in Annex D.

Initial configuration

DUT is configured for int8 and TOPS/watt measurement.

DUT is Switched OFF.

Power meter is Switched ON and connected to the DUT.

Test procedure

Step	Test procedure	Expected result
1	If audit is required, review AI Application source code to ensure unbiased implementation of int8 TOPS testing.	Software source code audit report for int8 TOPS testing.
2	Switch DUT ON, adjust the screen brightness to the lowest level, turn OFF the Bluetooth, turn OFF all notifications and turn ON the flight mode.	DUT is ON, and is in flight mode with all radios (e.g., cellular radio, BT, Wi-Fi etc) & notifications turned OFF.
3	Record the current and voltage.	The current curve and the voltage are displayed.
4	Wait until the current is stable, i.e. the current curve is stable [+/-5%].	The current is stable.
5	Record the background current and the voltage for 60 seconds, compute the average value.	The value of average background current and average voltage are obtained.
6	Run the test scripts for int8 Test Model, record the inference time and compute the average inference current.	The inference time and the average inference current value are obtained.
7	Compute int8 TOPS and compare the result with the value specified in the requirement TS.47_3.1_REQ_001.	The int8 TOPS result meets requirement TS.47_3.1_REQ_001.
8	Compute int8 TOPS/Watt and compare the result with the value specified in the requirement TS.47_3.1_REQ_003.	The int8 TOPS/Watt result meets requirement TS.47_3.1_REQ_003.

3.2 Testing floating point implementation

Test purpose

To verify the DUT can meet the minimum requirements of float16 TOPS and float16 TOPS/w.

Referenced requirements

TS47_3.1_REQ_002	An AI Mobile Device SHOULD have a minimum of (0.5) float16 TOPS.
TS47_3.1_REQ_004	An AI Mobile Device SHOULD have a minimum of (0.3) float16 TOPS/Watt.

Preconditions

The output of converted model should closely match the output of reference model based on proposed approach in the Annex E.

Scripts to pre-process the test dataset, run the test model and measure TOPS.

AI Application for TOPS testing shall have the following characteristics:

a) Source code must be available for software auditing, and
b) AI application shall support the NN baseline model used to evaluate TOPS performance, and

c) AI application shall create a test report with TOPS configuration and performance.

Test Dataset

Test dataset shall be publicly available. Example of public datasets provided in Annex D.

Initial configuration

DUT is configured for float16 TOPS and float16 TOPS/watt measurement.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

DUT is Switched OFF.

Power meter is Switched ON and connected to the DUT.

Test procedure

Step	Test procedure	Expected result
1	If audit is required, review AI Application source code to ensure unbiased implementation of float32 TOPS testing.	Software source code audit report for float32 TOPS testing.
2	Switch DUT ON, adjust the screen brightness to the lowest level, turn OFF the Bluetooth, turn OFF all notifications and turn ON the flight mode.	DUT is ON and is in flight mode with all radios (e.g., cellular radio, BT, Wi-Fi) and notifications turned OFF.
3	Record the current and voltage.	The current curve and the voltage are displayed.
4	Wait until the current is stable, i.e., the current curve is stable [+/-5%].	The current is stable.
5	Record the background current and the voltage for 60 seconds, compute the average value.	The value of average background current and average voltage are obtained.
6	Run the test scripts for float16 Test Model, record the inference time and compute the average inference current.	The inference time and the average inference current value are obtained.
7	Compute float16 TOPS and compare the result with the value specified in the requirement TS.47_3.1_REQ_002.	The float16 TOPS result meets requirement TS.47_3.1_REQ_002.
8	Compute float16 TOPS/Watt and compare the result with the value specified in the requirement TS.47_3.1_REQ_004.	The float16 TOPS/Watt result meets requirement TS.47_3.1_REQ_004.

4 Software Functions

4.1 Deep learning model update

Test purpose

To verify that the DUT can support deep learning model update.

Referenced Requirements

TS47_3.2_REQ_001	An AI Mobile Device SHALL support on-device model updates of an existing deep learning network.

Preconditions

- **Updated Model preparation**

  1. Check the AI models that the DUT already supports (Applicability Table in Annex A)
  2. Pick one of the models and modified its weights to create the Updated Model.

- **Dataset preparation**

  Prepare a dataset that matches the selected model for inferencing usage.

Initial configuration

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	Switch DUT on.	DUT is in idle mode.
2	Load the updated model on DUT.	The updated model is successfully loaded with no exception.
3	Process the dataset using the updated model.	The updated model can be run on the device successfully and can be used for inferencing.

4.2 Native API requirements

Test purpose

To verify that the DUT has native APIs to expose AI hardware functions.

Referenced Requirements

TS47_3.2_REQ_002	An AI Mobile Device SHALL support native APIs to expose the AI hardware functions.

Preconditions

Use an AI Application for the DUT modified to use native APIs and the Source Code available for review. For modification of the AI application, to use native APIs, use OEM supplied SDK and documentation.

Reference input data and output showing the list of native APIs expected to be used.

Initial configuration

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	If audit is required, review AI Application source code to ensure support for Native APIs that expose AI hardware functions.	Software source code audit report for Native API support.
2	Switch DUT ON.	DUT is ON without network connection.
3	Load AI Application adapted to use native APIs.	AI Application is loaded.
4	Run the AI Application adapted to use native API on DUT..	AI Application runs successfully.
6	Compare reference output with actual output result.	Actual output shows the expected native APIs were called.

4.3 Application APIs requirements

Test purpose

To verify that DUT provides application APIs for the following commonly used AI application categories:

1. Computer Vision (CV),
2. Automatic Speech Recognition (ASR),
3. Natural Language Understanding (NLU) models.

Referenced Requirements

TS47_3.2_REQ_003	An AI Mobile Device SHALL support application APIs (See Appendix A) for native and third-party applications to access Computer Vision (CV), Automatic Speech Recognition (ASR), Natural Language Understanding (NLU) models.

Preconditions

Use AI Application for the DUT that is modified to use native APIs for the following AI applications:

- Computer Vision (CV)
- Automatic Speech Recognition (ASR)
- Natural Language Understanding (NLU)

Note: For modification of selected AI application(s) to use native APIs use OEM supplied SDK and documentation.

For the selected AI application(s), source code is available for review.

Prepare the reference input dataset and the expected list of native application APIs to be called.

Initial configuration

DUT’s network connection (e.g., Wi-Fi, cellular) is Switched OFF.

DUT is loaded with application API test script.

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	If audit is required, review each Application API source code to ensure support for: • Computer Vision (CV) • Automatic Speech Recognition ASR) • Natural Language Understanding (NLU)	Source code audit report with API support results.
2	Switch DUT ON.	DUT is in ON without network connection.
3	Load Computer Vision application adapted to use native API application.	Computer Vision application is loaded successfully.
4	Run the loaded Computer Vision application on the DUT with the reference input dataset.	The expected native APIs are invoked successfully.
5	Load Automatic Speech Recognition application adapted to use native API application.	Automatic Speech Recognition application loaded successfully.
6	Run the loaded Automatic Speech Recognition application on the DUT with the reference input dataset.	The expected native APIs are invoked successfully.
7	Load Natural Language Understanding application adapted to use native API application.	Natural Language Understanding application loaded successfully.
8	Run Natural Language Understanding application on the DUT with the reference input dataset.	The expected native APIs are invoked successfully.

4.4 Model Format conversion

Test purpose

To verify that DUT has the SDK to convert model format to its native format so that the model can be run successfully on the DUT.

Referenced Requirements

TS47_3.2_REQ_004	An AI Mobile Device SHOULD provide an SDK to convert DNN models from an existing format to the native format of the AI mobile device. Non-exhaustive examples of DNN model file format are: *.ckpt or *.pb, *.tflite, *.prototxt, *.pb or *.pth or *.pt, *.json and *.onnx.

Preconditions

OEM provides list of NN model formats that can be converted to DUT's native format.

OEM provides SDK to convert the format of the test model to native format and SDK shall be available with the following characteristics:

- OEM SDK shall be available for technical evaluation
- OEM SDK shall support conversion of the model format to the Native format of the DUT

Select a NN model format supported by OEM's SDK and prepare NN model in this format.

Initial configuration

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	Run the SDK to convert the format of test model to native format.	Test model is successfully converted to a DUT supported format.
2	Switch DUT ON.	DUT is in idle mode.
3	Load the converted model on DUT.	The converted model is successfully loaded with no exception.

4.5 Customized Operator

Test purpose

To verify that DUT support new operator customization.

Referenced Requirements

TS47_3.2_REQ_005	An AI Mobile Device SHOULD provide an SDK to support definition of new customized Deep Learning operators.

Preconditions

Select an AI Application that runs on DUT with the following characteristics:

- Source code is available for review
- Supports use of customized Deep Learning operator(s)
- Supports baseline NN model
- Supports modified NN model using customized Deep Learning operator(s)
- Generates an output for the baseline and custom NN models

OEM supplies SDK for creation of custom operators.

Create a NN Model that utilizes customised operator(s).

Prepare input test dataset.

Prepare reference output result with customised Deep Learning operator(s).

Initial configuration

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	Use OEM SDK to custom operator.	Custom operator created successfully.
2	Created a NN model utilizing the custom operator.	NN model utilizing the custom operator successful created.
3	Switch DUT on.	DUT is in idle mode.
4	Load the NN model utilizing the custom operator on DUT.	The NN model utilizing the custom operator is successfully loaded with no exception.
5	Run the application with NN model utilizing the custom operator with the reference input dataset	Output result with test dataset with NN model utilizing custom operator(s) is as expected.

5 Inference Performance

Note: Section for inference test cases when requirements are defined.

6 AI Application Requirements

6.1 Biometric Performance

Execute biometric performance tests specified in either section 6.1.1 or 6.2.1

6.1.1 Biometric Performance – without FIDO certification

This section defines test cases for the various biometric requirements associated with specific performance metrics when FIDO certification is not available.

6.1.1.1 Common Test Procedures for Biometric Performance

This section defines a common procedure to perform Biometric performance testing for 2D, 3D and fingerprint recognition.

Preconditions

As required by FIDO Biometrics Requirements [3], chapter 5.

Initial configuration

As required by FIDO Biometrics Requirements [3], chapter 5.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	Switch the DUT ON and lock the screen.	The screen is locked.
2	FAR and FRR test procedure for Device Unlock, Application Login & Payment	Mean of FAR and FRR are obtained.

Step	Test procedure	Expected result
	Authorization refer to FIDO Biometrics Requirements [3], 5.1.3.	
3	Check the result.	Both FAR and FRR results meet the relevant requirement.

6.1.1.2 2D facial biometric system performance

Test purpose

To verify that DUT meets the 2D facial biometric performance requirements.

Referenced requirements

TS47_3.4.1_REQ_001	An AI Mobile Device SHOULD support a 2D facial biometric system.
TS47_3.4.1_REQ_004	An AI Mobile Device supporting 2D facial biometric system SHALL support the biometric KPI requirement TS47_3.4.1_REQ_004.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization.
TS47_3.4.1_REQ_004.1	2D Facial FAR <= (0.002)% and FRR <= (3)% simultaneously

Preconditions

As defined in section 6.1.1.

Initial configuration

As defined in section 6.1.1.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

Test procedure

As defined in section 6.1.1, with FAR and FRR performance as defined for 2D facial biometric in section 6.1.2.

6.1.1.3 3D facial biometric system performance

Test purpose

To verify that DUT meets the 3D facial biometric performance requirements.

Referenced requirements

TS47_3.4.1_REQ_002	An AI Mobile Device SHOULD support a 3D facial biometric system.
TS47_3.4.1_REQ_005	An AI Mobile Device supporting 3D facial biometric system SHALL support the biometric KPI requirement TS47_3.4.1_REQ_005.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization.

TS47_3.4.1_REQ_005.1	3D Facial FAR <= (0.001)% and FRR <= (3)% simultaneously.

Preconditions

As defined in section 6.1.1.

Initial configuration

As defined in section 6.1.1.

Test procedure

As defined in section 6.1.1, with FAR and FRR performance as defined for 3D facial biometric in section 6.1.3.

6.1.1.4 Fingerprint biometric system performance

Test purpose

To verify that DUT meets the fingerprint biometric performance requirements.

Referenced requirements

TS47_3.4.1_REQ_003	An AI Mobile Device SHOULD support a fingerprint biometric system.
TS47_3.4.1_REQ_006	An AI Mobile Device supporting fingerprint biometric system SHALL support the biometric KPI requirement TS47_3.4.1_REQ_006.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization.
TS47_3.4.1_REQ_006.1	Fingerprint FAR <= (0.002)% and FRR <= (3)% simultaneously.

Preconditions

As defined in section 6.1.1.

Initial configuration

As defined in section 6.1.1.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

Test procedure

As defined in section 6.1.1 with FAR and FRR performance as defined for fingerprint biometric in section 6.1.4.

6.1.2 Biometric Performance – with FIDO certification

This section defines test cases for the various biometric requirements when FIDO certification is available.

6.1.2.1 2D facial biometric system performance

Referenced requirements

TS47_3.4.1_REQ_001	An AI Mobile Device SHOULD support a 2D facial biometric system.

TS47_3.4.1_REQ_004	An AI Mobile Device supporting 2D facial biometric system SHALL support the biometric KPI requirement TS47_3.4.1_REQ_004.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization.
TS47_3.4.1_REQ_004.1	2D Facial FAR <= (0.002)% and FRR <= (3)% simultaneously
TS47_3.4.1_REQ_007	The biometric key performance indicators (KPIs) for the supported biometric system SHOULD be certified by one or more of the following programs: Fast IDentity Online (FIDO) Alliance Biometric Component Certification Program. Internet Finance Authentication Alliance (IFAA) biometric Certification Program.

Test purpose

To verify that DUT meets the 2D facial biometric performance requirements.

Preconditions

OEM provided FIDO certification for 2D facial biometric performance.

Initial configuration

None.

Test procedure

Step	Test procedure	Expected result
1	Check FIDO/IFFA certification for 2D facial recognition.	OEM provided FIDO certification stating 2D facial recognition requirements defined above are met.

6.1.2.2 3D facial biometric system performance

Test purpose

To verify that DUT meets the 3D facial biometric performance requirements.
Referenced requirements

TS47_3.4.1_REQ_002	An AI Mobile Device SHOULD support a 3D facial biometric system.
TS47_3.4.1_REQ_005	An AI Mobile Device supporting 3D facial biometric system SHALL support the biometric KPI requirement TS47_3.4.1_REQ_005.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization.
TS47_3.4.1_REQ_005.1	3D Facial FAR <= (0.001)% and FRR <= (3)% simultaneously.

TS47_3.4.1_REQ_007	The biometric key performance indicators (KPIs) for the supported biometric system SHOULD be certified by one or more of the following programs: Fast IDentity Online (FIDO) Alliance Biometric Component Certification Program. Internet Finance Authentication Alliance (IFAA) biometric Certification Program.

Preconditions

OEM provided FIDO certification for 2D facial biometric performance.

Initial configuration

None.

Test procedure

Step	Test procedure	Expected result
1	Check FIDO certification for 3D facial recognition.	OEM provided FIDO certification stating 3D facial recognition requirements defined above are met.

6.1.2.3 Fingerprint biometric system performance

Test purpose

To verify that DUT meets the fingerprint biometric performance requirements.

Referenced requirements

TS47_3.4.1_REQ_003	An AI Mobile Device SHOULD support a fingerprint biometric system.
TS47_3.4.1_REQ_006	An AI Mobile Device supporting fingerprint biometric system SHALL support the biometric KPI requirement TS47_3.4.1_REQ_006.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization.
TS47_3.4.1_REQ_006.1	Fingerprint FAR <= (0.002)% and FRR <= (3)% simultaneously.
TS47_3.4.1_REQ_007	The biometric key performance indicators (KPIs) for the supported biometric system SHOULD be certified by one or more of the following programs: Fast IDentity Online (FIDO) Alliance Biometric Component Certification Program. Internet Finance Authentication Alliance (IFAA) biometric Certification Program.

Preconditions

OEM provided FIDO certification for fingerprint performance.

Initial configuration

None.

Test procedure

Step	Test procedure	Expected result
1	Check FIDO certification for fingerprint recognition.	OEM provided FIDO certification stating fingerprint recognition requirements defined above are met.

6.2 On-device Image Processing

6.2.1 Photo scene detection

Test purpose

To verify that DUT meets the photo scene detection and recognition requirements.

Referenced requirements

TS47_3.4.2_REQ_001	An AI Mobile Device SHOULD have optical character recognition (OCR) capability on the device.
TS47_3.4.2_REQ_002	An AI Mobile Device SHOULD have image detection, image classification and image segmentation capabilities on the device.
TS47_3.4.2.1_REQ_001	The AI Mobile Device SHOULD support photo scene detection and recognition where the User has the ability to consent to their use.
TS47_3.4.2.1_REQ_001.1	If REQ_001 is supported then the AI Mobile Device SHALL support Identification of one or more objects in different scenes such as portraits, landscapes, foods, night scenes and texts, etc.
TS47_3.4.2.1_REQ_001.2	If TS47_3.4.2.1 REQ_001 is supported then the AI Mobile Device SHALL support Scene detection capabilities to optimize camera settings for image capture based on scene content.

Preconditions

OEM provides list of scene categories supported by the DUT e.g., portrait, landscape, food, night scene, text, pets, flowers etc.

Set of images covering all scene categories supported by the DUT e.g., portrait, landscape, food, night scene, text, pets, flowers etc.

Application to exercise the photo categorization.

Initial configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

The photo scene detection and recognition function on DUT is enabled with user consent.

Test procedure

Step	Test procedure	Expected result
1	Switch the DUT camera ON.	The DUT's camera is ON.
2	Take photo of a randomly selected image from the test dataset.	The DUT identifies photo correctly as either a portrait, a landscape, a food, a night scene or a text. The photos are saved in an album on the DUT.
3	Repeat step 2 with the remaining photos in the test dataset.	

6.2.2 Photo categorisation

Test purpose

To verify that DUT meets the photo scene detection and recognition requirements.

Referenced requirements

TS47_3.4.2_REQ_003	An AI Mobile Device SHOULD have face detection and face clustering capabilities within a group of photos on the device.
TS47_3.4.2.1_REQ_007	The AI Mobile Device SHOULD support automatic classification of photos in an album by different categories.

Preconditions

OEM provides the list of categories supported by the DUT.

Set of images covering all categories supported by the DUT e.g., portraits, landscape, food, party, pets, flowers etc.

Initial configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

The photo scene detection and recognition function on DUT is enabled with user consent.

Test procedure

Step	Test procedure	Expected result
1	Switch the DUT camera ON.	The DUT's camera is ON.
2	Using DUT, take photos of different scenes supported by the DUT e.g., portraits, landscapes, foods, etc.	The photos are saved in an album on the DUT according to the defined categories.

6.2.3 Text detection and recognition

Test purpose

To verify that DUT supports automatic language detection, text detection and recognition.

Referenced requirements

TS47_3.4.2.1_REQ_002	The AI Mobile Device SHOULD support text detection and recognition of installed language packages, where the User has the ability to consent to the text detection and recognition use.
TS47_3.4.2.1_REQ_003	The AI Mobile Device SHOULD support automatic language detection.

Preconditions

Images containing text in language(s) supported by the DUT.

Initial configuration

DUT is Switched ON.

DUT's network connection(e.g., Wi-Fi, cellular) is Switched OFF.

Text detection and recognition function is enabled with user consent.

Test procedure

Step	Test procedure	Expected result
1	Apply text detection on text image dataset.	Text in the picture identified. Language of the text in the picture identified.

6.3 Voice Processing

6.3.1 Test Environment Preparation for Speech Test

It is recommended to carry out tests in an environment where the conditions can be controlled, e.g., in a reverberation chamber, and the environment has following characteristics.

- The environment reverberation time is between 0.4 second and 0.7 seconds.
- The environment frequency is between 100Hz and 8kHz.
- The background noise scenarios are defined in Scenario 1-3 below.
- The test subjects are defined in Scenario 4-7 below.
- The distance between DUT and test subject is between [30] cm and [60] cm.

- Other specifications for the reverberation chamber can take reference from ETSI EG 202 396-1 (part 6). [8]

The voice source in Figure 1 is shown as a human but it can be an electronic speaker.

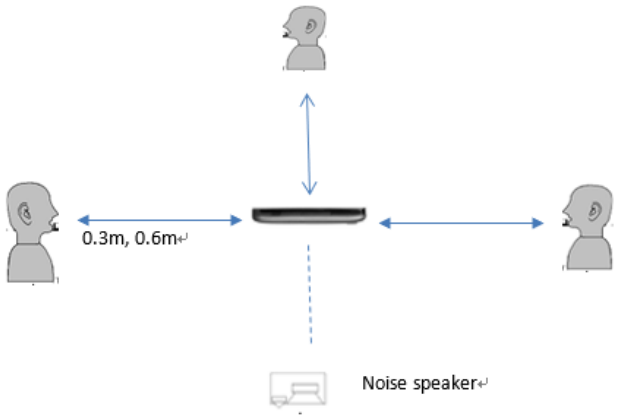

Figure 1 the AI speech recognition system

The background noise definition:

Noise Scenario No.	Noise Scenario Category	Distance between noise source & DUT	Noise source level
1	Background with low noise	1.5m	Less than 40dB
2	Background with medium noise	1.5m	40-60dB
3	Background with high noise	1.5m	60-70dB

1. The background noise source should come from a TV show, a spoken conversation, music and a radio broadcast.
2. The noise source is recommended to use the default language configured on the device.
3. The noise should be:

   3 minutes TV show

   3 minutes human spoken conversation

   3 minutes music (CD or mp3 file)

   3 minutes radio broadcast (anything)

   The background noise can be looped as required

The voice source speed/frequency/tone:

Scenario No.	Gender distribution	Age of the tester	Proportion	Number of People
4	Female 50%	Below 7	20%	For manual testing: More than [20] people. For tool testing: [50-100] people.
5	Male 50%	7-18	20%	
6		19-50	40%	
7		Beyond 50	20%	

6.3.2 Automatic Speech Recognition Capabilities

Test Purpose

To verify that DUT meets the automatic speech recognition requirements.

Referenced Requirements

TS47_3.4.3_REQ_001	The AI Mobile Device SHOULD have speech ability.
TS47_3.4.3_REQ_002	The AI Mobile Device SHOULD support Automatic speech recognition (ASR) capabilities where the User has the ability to consent to ASR.
TS47_3.4.3_REQ_005	If the AI Mobile Device supports Voice Assistant then the requirements in section 3.4.3.1 SHALL apply.
TS47_3.4.3.1_REQ_001	AI Mobile Device SHALL support the Automatic speech recognition (ASR) capabilities.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1.

If the DUT supports Voice Assistant or the DUT has an in-built/pre-loaded ASR application, OEM provides the necessary information needed to use the application e.g., list of words or command supported by the ASR application, how to activate/enable the ASR application.

Otherwise, load automatic speech recognition application e.g., speech-to-text application, voice command application. If this is not possible then the test is not applicable.

Test Dataset

Record a set of words of all the test subjects with the DUT’s configured default language.

The target voice volume should be more than 70dB.

The number of recorded words should be repeated for [20] times each.

The gender and age distribution of test subject refer to Scenario 4-7 in Section 6.3.1.

Initial Configuration

DUT is Switched ON.

DUT’s network connection (e.g., Wi-Fi, cellular) Switched OFF..

Test Procedure

Step	Test procedure	Expected result
1	Play the randomly selected recorded word one by one under each of the background noise Scenario 1-3.	The output from the ASR application is as expected, e.g., for speech-to-text application the text represents the spoken word(s)..

6.3.3 Natural Language Understanding Capabilities

Test Purpose

To verify that DUT meets the requirement for Natural Language Understanding.

Referenced Requirements

TS47_3.4.3_REQ_001	The AI Mobile Device SHOULD have speech ability.
TS47_3.4.3_REQ_003	The AI Mobile Device SHOULD support Natural Language Understanding (NLU) capabilities where the User has the ability to consent to NLU.
TS47_3.4.3_REQ_005	If the AI Mobile Device supports Voice Assistant then the requirements in section 3.4.3.1 SHALL apply.
TS47_3.4.3.1_REQ_001.1	AI Mobile Device SHALL support the Natural Language Understanding (NLU) capabilities.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1.

NLU dataset: conversation scenarios for a single subject.

If the DUT supports Voice Assistant or the DUT has an in-built/pre-loaded NLU application, OEM provides the necessary information needed to use the application e.g., how to activate/enable the NLU application.

Otherwise, load NLU application. If this is not possible then the test is not applicable.

The dialogs can be recorded in a quiet environment described in Scenario 1 in Section 6.3.1 by a test subject described in Scenario 6 in Section 6.3.1. The test dataset should include at least 1 of the following proposed categories in the DUT's default language:

Category	Dialogs examples
Send messages:	"Open message" "Read the latest message" "Reply to this number" "Attach one photo to this message" "Send this message".

Operate browser	"Open browser and search where is the capital of France" "Open browser and search how many kinds of cats are there" "Open browser and search what is nuclear" "Open browser and search what are the main types of roses" "Open browser and search why is the sky blue"
Call for location service	"Find the nearest restaurant" "Call restaurant" i.e., call the restaurant search recommended. "How to get there?" i.e., how to get to the searched venue.

Initial Configurations

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF unless network required to carry out the command.

Test Procedure

Step	Test procedure	Expected result
2	The NLU application is activated, randomly play each of the recorded NLU phrase from the dataset under background noise Scenario 1-3. Determine whether the voice assistant responds correctly.	The voice assistant responds correctly.

6.3.4 Text to Speech Capabilities

Test Purpose

To verify that DUT can meet the text-to-speech requirement.

Referenced Requirements

TS47_3.4.2_REQ_001	An AI Mobile Device SHOULD have optical character recognition (OCR) capability on the device.
TS47_3.4.3_REQ_001	The AI Mobile Device SHOULD have speech ability.
TS47_3.4.3_REQ_004	The AI Mobile Device SHOULD support Synthesized Voice (Text-To-Speech (TTS) capabilities where the User has the ability to consent to TTS.
TS47_3.4.3_REQ_005	If the AI Mobile Device supports Voice Assistant then the requirements in section 3.4.3.1 SHALL apply.
TS47_3.4.3.1_REQ_001.2	AI Mobile Device SHALL support the Synthesized Voice (Text-To-Speech (TTS)) capabilities.

Preconditions

If the DUT supports Voice Assistant or the DUT has an in-built/pre-loaded TTS application, OEM provides the necessary information needed to use the application.

Otherwise, load TTS application. If this is not possible then the test is not applicable.

Test Dataset

Prepare at least [5] different text sentences, and can be messages/ e-mails / handwritten text that contain numbers, letters, and symbols.

Initial Configurations

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

The default language configured on the device is recommended to be used for testing.

Test Procedure

Step	Test procedure	Expected result
1	The TTS application is activated, TTS dataset are sent to the DUT in a text message and get the TTS assistant to readout the received text, and check whether the readout matches the text.	The voice assistant speaks the message content correctly.

6.3.5 Voice Assistant Functions

6.3.5.1 Voiceprint Recognition Performance - quiet environment

Test Purpose

To verify that DUT meets the voice trigger performance requirements.

Referenced Requirements

TS47_3.4.3.1_REQ_002	The AI Mobile Device SHALL support voice trigger, and its specific requirements are listed in the following sub requirements: TS47_3.4.3.1_REQ_002.1, 002.2 and_002.3
TS47_3.4.3.1_REQ_002.1	The AI Mobile Device SHOULD support voiceprint recognition for preventing people other than the device's owner from triggering voice assistant.
TS47_3.4.3.1_REQ_002.2	In a quiet environment, the following SHALL be required: The true acceptance rate (TAR) >= (90)%, and the false acceptance rate (FAR) of voiceprint recognition <= (20)%.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1 and background noise scenarios as defined in Scenario 1.

Test Dataset

Record wake-up words of all the test subjects.

The target voice volume should be more than 70dB.

The number of the recorded wake-up words should be repeated for [20] times.

The gender and age distribution of test subject refer to Scenario 4-7 in Section 6.3.1.

Initial Configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

Test subject is labelled as v(i), where i = 1, 2, .. n, where n is the total number of the test subjects.

Test Procedure

Step	Test procedure	Expected result
1	Switch on the voice assistant.	The voice assistant is on.
2	Conduct the test in quiet environment, for each v(i) , enroll its wake-up word according to the system prompts.	Wake-up word setting is successful.
3	Play the v(i) recorded wake-up word and record the success wake-up times as M(i).	The value of M(i) are obtained.
4	Play the rest (n-1) test subject's wake-up word, check whether the voice assistant is woken-up and record the number of success as N(i).	The values of N(i) are obtained.
5	Compute TAR and FAR of voiceprint recognition, compare the result with the value specified in the requirement TS47_3.4.3.1_REQ_002.2. TAR=$\frac{\sum_{i=1}^{n} M(i)}{n*20}$, FAR of voiceprint recognition=$\frac{\sum_{i=1}^{n} N(i)}{n*(n-1)}$	The TAR and FAR of voiceprint recognition result meets requirement TS47_3.4.3.1_REQ_002.2.

6.3.5.2 Voiceprint recognition performance – noisy environment

Test purpose

To verify that DUT meets the voice trigger performance requirements.

Referenced requirements

TS47_3.4.3.1_REQ_002	The AI Mobile Device SHALL support voice trigger, and its specific requirements are listed in the following sub requirements: TS47_3.4.3.1_REQ_002.1, (002.2) and_002.3
TS47_3.4.3.1_REQ_002.1	The AI Mobile Device SHOULD support voiceprint recognition for preventing people other than the device's owner from triggering voice assistant.
TS47_3.4.3.1_REQ_002.3	In a noisy environment, the following SHALL be required: TAR >=(80)%, and FAR of voiceprint recognition <= (20)%.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1 and background noise scenarios as defined in Scenario 3.

Test Dataset:

Record wake-up words of all the test subjects.

The target voice volume should be more than 70dB.

The number of the recorded wake-up words should be repeated for [20] times.

The gender and age distribution of test subject refer to Scenario 4-7 in Section 6.3.1.

Initial configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

Test subject is labelled as v(i), where i = 1, 2, …, n, where n is the total number of the test subjects.

Test procedure

Step	Test procedure	Expected result
1	As defined in section 6.4.5.2.2 with noise environment as defined in Scenario 23 in section 6.4.1.	The TAR and FAR of voiceprint recognition result meets requirement TS47_3.4.3.1_REQ_002.3.

6.3.5.3 On-device Speech Recognition

Test Purpose

To verify that the voice assistant of DUT has the local capability to change system setting even in the case that without network connection and invoke native application.

Referenced Requirements

TS47_3.4.3.1_REQ_003	The AI Mobile Device SHALL have on-device speech recognition library (i.e. with no access to the Internet) for changing the system setting (e.g. Turn Bluetooth on/off via voice assistant) and invoking the native applications (e.g. send SMS via voice assistant).

Preconditions

Test environment and test subject preparation as required in Section 6.3.1.

Test Dataset

The dialogs should be recorded in a quiet environment described in Scenario 1 in Section 6.3.1 by test subjects in Scenario 6 in Section 6.3.1.

Each test subject should record 5 different commands such as those shown below:

Command examples	
Set the volume of the phone	Calculate mathematics
Play local music	Turn on the flashlight
Set alarms on the phone	Launch Applications
Open the Gallery	Add memos

Initial Configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

Test Procedure

Step	Test procedure	Expected result
1	Play a random command from the test dataset and check whether voice assistant responds to the command correctly.	The voice assistant responds or executes the command correctly.
2	Repeat 4 times for different command from the test dataset.	The voice assistant responds or execute the command correctly.

6.3.5.4 Interaction with Third-party Applications

Test Purpose

To verify that voice assistant on DUT supports interaction with third-party applications.

Referenced Requirements

TS47_3.4.3.1_REQ_004	The AI Mobile Device SHOULD have access to different categories of applications and invoke these applications' services and functions via voice assistant.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1

Test Dataset

The dialogs in test dataset should be recorded in a quiet environment described in Scenario 1 in Section 6.3.1 by test subjects in Scenario 6 in Section 6.3.1.

OEM provides the list of commands supported by the DUT.

Each test subject should record [2] commands.

Initial Configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF unless network required to carry out the command.

Test Procedure

Step	Test procedure	Expected result
1	Randomly play a command from the test dataset and check whether voice assistant responds correctly to the command.	The voice assistant responds or executes the command correctly.
2	Repeat once with a different command from the test dataset.	The voice assistant responds or executes the command correctly.

6.3.5.5 Information Search

Test Purpose

To verify that voice assistant on DUT supports information search.

Referenced Requirements

TS47_3.4.3.1_REQ_005	The AI Mobile Device SHALL support information search by on-device voice assistant.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1.

Test Dataset

The command in the test dataset should be recorded in a quiet environment described in Scenario 1 in Section 6.4.1 by test subjects in Scenario 6 in Section 6.3.1.

Each test subject should record [5] of the following proposed categories,

Category	Command Examples
Business & finance	What is the stock market for GOOG?
Food & drink	How to cook t beef steaks
Games & fun	Play Hip-hop music
Health & fitness	What is the benefit of doing exercise?
Children & family	Where is the tallest building in the world?
Navigation service	How do I get to the nearest petrol station?
Movies, Music, photos	Open the photo that was taken in Beijing last month
News & books	Display the news for yesterday
Productivity	Is there any arrangement on 19th of March?
Shopping	What is the price for iPhone X on Amazon?
Social & Communications	Check my tweets on twitter
Sports	What is the UK premier league football match for tonight

Category	Command Examples
Travel & transportation	List some hotels in Dubai
Weather	Will it rain tomorrow?
Knowledge & encyclopedia	Give me an introduction to the Law of Gravity

Initial Configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched ON.

Test Procedure

Step	Test procedure	Expected result
1	Randomly play a command from the test dataset and check whether voice assistant responds correctly.	The voice assistant responds or executes the command correctly.
2	Repeat 4 times for different command from the test dataset.	The voice assistant responds or executes the command correctly.

6.3.5.6 Interaction with Smart Devices

Test Purpose

To verify that voice assistant on DUT supports interaction with smart devices.

Referenced Requirements

TS47_3.4.3.1_REQ_006	The AI Mobile Device SHOULD support interaction with smart devices (e.g. home appliances) via voice assistant.

Preconditions

Test environment and test subject preparation as required in Section 6.3.1.

Test Dataset

The dialog should be recorded in a quiet environment described in Scenario 1 in Section 6.3.1 by test subjects in Scenario 6 in Section 6.3.1.

One example is turning smart device on/off and more dialogs can be included if needed.

OEM provides list of supported smart devices.

Initial Configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched ON.

Test Procedure

Step	Test procedure	Expected result
1	Play the command from the test dataset and check whether voice assistant interacts with smart device successfully.	The voice assistant interacts with the smart device successfully.

6.4 Augmented Reality (AR)

6.4.1 AI capabilities for AR applications

Test Purpose

To verify that DUT provides AI capabilities for AR native and third-party applications.

Referenced Requirements

TS47_3.4.4_REQ_001	The AI Mobile Device SHOULD provide the following AI capabilities for AR native and third-party applications: 1. Hand gesture recognition. 2. Hand skeleton tracking. 3. Human body pose recognition. 4. Human body skeleton tracking.

Preconditions

OEM provides list of AR application categories supported by the DUT.

Initial Configuration

DUT is Switched ON.

DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

The setting for camera access for DUT is turned on.

Test procedure

Step	Test procedure	Expected result
1	Open the first AR application, perform gesture in front of the camera, and check whether appropriate gesture is depicted correctly on the screen	Screen shows appropriate gesture correctly.
2	Repeat step 1 for the next supported AR human pose.	Screen shows the results correctly according to the AR human pose.

6.4.2 AR Emoji application

Test purpose

To verify that DUT supports AR emoji functions.

Referenced requirements

TS47_3.4.4_REQ_002	1. The AI Mobile Device SHOULD support the following AR Emoji operations a. Creating customized AR-based Emoji. b. Tracking User's facial movement and expression and render these on the AR-based Emoji.

Preconditions

Any emoji required by the test loaded to the device.

Initial configuration

DUT is Switched ON.DUT's network connection (e.g., Wi-Fi, cellular) is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	Enable AR emoji function.	The AR emoji function is enabled.
2	Create/select an AR emoji. Request test subject to make various movements supported by the emoji e.g., move head to left/right, blink eyes, open/close mouth, make sad/happy face, etc. Check Emoji follows the facial expressions.	The AR emoji does the same actions as the test subject.

7 Privacy and Security

7.1 Privacy

7.1.1 AI Application Personal Data Processing Control

See Annex B

7.1.2 FPE AI Application Default Setting

See Annex B

7.2 Security

7.2.1 Requirement of personal data collection control

See Annex C

7.2.2 Requirement of Off toggle switches

Test purpose

To verify whether there are Off ‘toggle’ switches that can be used to turn OFF the functionality, except as permitted or required by applicable law.

Referenced requirements

TS47_4.2_REQ_003	Off ‘toggle’ switches SHALL turn off the functionality, except as permitted or required by applicable law.

Preconditions

DUT reset to factory default state.

OEM provides list all the Off ‘toggle’ switches.

Initial configuration

None.

Test procedure

Step	Test procedure	Expected result
1	Check all OFF ‘toggle’ switches provided by OEM do turn OFF the corresponding AI function.	All the OFF ‘toggle’ switches turn OFF the functionality, except as permitted or required by applicable law.

7.2.3 Security for AI applications

7.2.3.1 Requirement of Secured Environment

See Annex C

7.2.3.2 Requirements of Data Encryption and Key Storage

See Annex C

7.2.3.3 Requirement of Biometric Data for authentication

See Annex C

7.2.3.4 Requirements of Biometric Data Storage

See Annex C

7.2.3.5 Requirement of Biometric Data replacement

Test purpose

To verify that Biometric Data replacement meets the secure requirements.

Referenced requirements

TS47_4.2.1_REQ_009	If Users' Biometric Data is replaced, the previous Biometric Data before the replacement SHALL be deleted completely and permanently and not be recoverable by data rollback.
TS47_4.2.1_REQ_013	When the Voiceprint Data is permanently and completely deleted, it SHALL NOT be recoverable by data rollback

Preconditions

Select Biometric data to use i.e., fingerprint, 2D facial, 3D facial or voiceprint.

Prepare selected Biometric Data for User 1 and User 2.

Initial configuration

DUT is Switched OFF.

Biometric Data for User 1 is pre-stored on DUT with user's consent.

Test procedure

Step	Test procedure	Expected result
1	Switch DUT ON.	DUT is in idle mode.
2	Use selected Biometric Data for User 1 to login/unlock AI application(s).	The Biometric Data for User 1 can login/unlock AI application(s).
3	Replace Biometric Data for User 1 with Biometric Data for User 2.	Users' Biometric Data is updated and Biometric Data for User 1 is deleted.
4	Use Biometric Data for User 1 to login/unlock AI application(s).	AI applications cannot login / unlocked.
5	Use Biometric Data for User 2 to login/unlock AI application(s)	The Biometric Data for User 2 can login/unlock AI application(s).
6	Check whether there are Biometric Data rollback operation button(s) provided by OEM.	No Biometric Data rollback operation button(s)
7	Execute Backup operation.	Except for Biometric Data**,** data permitted by the user is backed up.
8	Delete Biometric Data for User 2.	Biometric Data for User 2 is deleted.
9	Execute Restore operation.	Except for Biometric Data**,** data that backed up in Step 7 is restored.
10	Use Biometric Data for User 2 to login/unlock AI application(s).	Biometric Data for User 2 cannot login/unlock AI application(s).

7.2.3.6 Requirement of device factory reset

Test purpose

To verify that the Biometric Data are wiped out and made unrecoverable by a device factory reset.

Referenced requirements

TS47_4.2.1_REQ_010	The Biometric Data SHALL be wiped and made unrecoverable by a device factory reset.
TS47_4.2.1_REQ_014	The Voiceprint Data SHALL be wiped and made unrecoverable by a device factory reset.

Preconditions

Select Biometric data to use i.e., fingerprint, 2D facial, 3D facial or voiceprint

Prepare Biometric Data for User 1.

Initial configuration

DUT is Switched OFF.

Biometric Data 1 is pre-stored on DUT with user's consent.

Test procedure

Step	Test procedure	Expected result
1	Switch DUT ON.	DUT is in idle mode.
2	Execute AI applications with Biometric Data for User 1.	AI applications can be executed with Biometric Data for User 1.
3	Execute factory reset on DUT	
4	Check the information of Biometric Data on DUT.	Biometric Data 1 is wiped out.
5	Execute AI applications with the Biometric Data for User 1.	AI applications cannot be executed.

7.2.3.7 Requirement of temporary Voiceprint Data

See Annex C

7.2.3.8 Requirement for Voice replay attack defence

Test purpose

To verify that the device can be resistant to voice replay attacks.

Referenced requirements

TS47_4.2.1_REQ_015	The device SHOULD be resistant to voice replay attacks.

Preconditions

Prepare voice replay samples.

OEM provides list of AI applications controlled with voiceprint.

The background noise is as defined for Scenario 1, 2 & 3 in section 6.3.1.

Voice replay test dataset should be prepared by recording the user's wake-up words for the AI application that support voice recognition.

Initial configuration

DUT is Switched OFF.

Test procedure

Step	Test procedure	Expected result
1	Switch DUT ON.	DUT is in idle mode.
2	Configure selected background noise scenario.	Background noise is active.
3	Open the AI applications that support voice recognition and use voice replay test dataset to perform voice replay attacks. under different background noise.	AI applications cannot execute with voice replay test dataset.
4	Repeat steps 2 to 3 for next background scenario listed in the Preconditions.	

7.2.4 Requirement for Data Processor transparency

See Annex C

7.2.5 Requirement for anti-user manipulation

See Annex C

Annex A Applicability Table

Applicability Table: A document, in the form of a questionnaire , which requires the device manufacturer to declare which requirement the device meets and to provide the necessary information for conducting tests.

TS.53 Test Cases Number	Test Case Applicability M = Mandatory O = Optional C = Conditional	Conditional Requirements
3.1 Integer implementation performance	**O**	
3.2 Testing floating point implementation	**O**	
4.1 Deep learning model update	**M**	
4.2 Native API requirements	**M**	
4.3 Application APIs requirements	**M**	
4.4 Model Format conversion	**O**	
4.5 Customized Operator	**O**	
6.1.1.2 2D facial biometric system performance – without FIDO certification	**C**	If 2D facial biometric system is supported then **M**
6.1.1.3 3D facial biometric system performance – without FIDO certification	**C**	If 3D facial biometric system is supported then **M**
6.1.1.4 Fingerprint biometric system performance – without FIDO certification	**C**	If Fingerprint biometric system is supported then **M**
6.1.2.1 2D facial biometric system performance - with FIDO certification	**C**	If 2D facial biometric system is supported then **M**
6.1.2.2 3D facial biometric system performance - with FIDO certification	**C**	If 3D facial biometric system is supported then **M**
6.1.2.3 Fingerprint biometric system performance - with FIDO certification	**C**	If Fingerprint biometric system is supported then **M**
6.2.1 Photo scene detection	**C**	This test case is **M** is any of the following are supported on the device: 1. Optical character recognition (OCR) 2. Image detection, image classification and image segmentation.

TS.53 Test Cases Number	Test Case Applicability M = Mandatory O = Optional C = Conditional	Conditional Requirements
		3. Photo scene detection and recognition where the User has the ability to consent to their use
6.2.2 Photo categorization	**O**	
6.2.3 Text detection and recognition	**O**	
6.3.2 Automatic Speech Recognition Capabilities	**C**	This test case is **M** is either of the following are supported on the device: 1. Speech ability 2. Automatic speech recognition (ASR) capabilities where the User has the ability to consent to ASR
6.3.3 Natural Language Understanding Capabilities	**C**	This test case is **M** is either of the following are supported on the device: 1. Speech ability 2. Natural Language Understanding (NLU)
6.3.4 Text to Speech Capabilities	**C**	This test case is **M** is either of the following are supported on the device: 1. Optical character recognition (OCR) 2. Speech ability 3. Synthesized Voice (Text-To-Speech (TTS)
6.3.5.1 Voiceprint Recognition Performance - quiet environment	**M**	
6.3.5.2 Voiceprint recognition performance – noisy environment	**M**	
6.3.5.3 On-device Speech Recognition	**M**	
6.3.5.4 Interaction with Third-party Applications	**O**	
6.3.5.5 Information Search	**M**	
6.3.5.6 Interaction with Smart Devices	**O**	
6.4.1 AI capabilities for AR applications	**O**	

TS.53 Test Cases Number	Test Case Applicability M = Mandatory O = Optional C = Conditional	Conditional Requirements
6.4.2 AR Emoji application	**O**	
7.1.1 AI Application Personal Data Processing Control	**M**	Letter in Annex B to be completed and signed / stamped
7.1.2 FPE AI Application Default Setting	**M**	Letter in Annex B to be completed and signed / stamped
7.2.1 Requirement of personal data collection control	**M**	Letter in Annex C to be completed and signed / stamped
7.2.2 Requirement of Off toggle switches	**M**	
7.2.3.1 Requirement of Secured Environment	**O**	Letter in Annex C to be completed and signed / stamped
7.2.3.2 Requirements of Data Encryption and Key Storage	**M**	Letter in Annex C to be completed and signed / stamped
7.2.3.3 Requirement of Biometric Data for authentication	**M**	Letter in Annex C to be completed and signed / stamped
7.2.3.4 Requirements of Biometric Data Storage	**M**	Letter in Annex C to be completed and signed / stamped
7.2.3.5 Requirement of Biometric Data replacement	**M**	
7.2.3.6 Requirement of device factory reset	**M**	
7.2.3.7 Requirement of temporary Voiceprint Data	**M**	Letter in Annex C to be completed and signed / stamped
7.2.3.8 Requirement for Voice replay attack defense	**O**	
7.2.4 Requirement for Data Processor transparency	**M**	Letter in Annex C to be completed and signed / stamped
7.2.5 Requirement for anti-user manipulation	**M**	Letter in Annex C to be completed and signed / stamped

Annex B Letter of Commitment One

_____(Company name) _____(DUT model) complies with these privacy requirements

TS.47 Requirement Number	Requirement	Is this requirement supported? Yes / No
TS47_4.1_REQ_003	AI Applications that process Personal Data SHALL be off by default unless processing exclusively takes place locally on the device.	
TS47_4.1_REQ_004	The AI Application on the AI Mobile Device SHALL be designed in such a way that a Data Processor will have the responsibility to: 1) Be transparent with the User on the nature of the input data used in the AI processing (e.g. personal files, biometrics, …). 2) Forbid transferring personal data processing off the device except if the User has explicitly agreed or other legal basis has been satisfied in accordance with the law. 3) Forbid transferring results of on-device AI processing containing personal data off the device except if the User has explicitly agreed or other legal basis has been satisfied in accordance with the law.	

_____(Company Representative Signature or Company Stamp)

_____(Company Representative Print Name)

_____(Company Representative Job Title)

_____(Date)

Annex C Letter of Commitment Two

_____(Company name)_____(DUT model) complies with these security requirements

TS.47 Requirement Number	Requirement	Is this requirement supported? Yes / No
TS47_3.4.2.1_REQ_006	The FPE functionality SHOULD be switched off by default.	
TS47_4.2_REQ_002	Except as required or permitted by applicable law, the User SHALL always remain in control of the collection of their personal data and its usage, in order to minimise the risk of malicious usage or data leakage.	
TS47_4.2_REQ_004	Techniques, such as 'Dark Patterns', that manipulate the User's choice SHALL NOT be used.	
TS47_4.2.1_REQ_003	Autonomous AI Mobile Device operations SHALL be controlled, and/or authorized by the authenticated User.	
TS47_4.2.1_REQ_005	Data and metadata for AI Mobile Device SHALL be stored with encryption with keys that are stored securely in a Secured Environment, e.g. Trusted Execution Environment (TEE).	
TS47_4.2.1_REQ_006	Biometric Data, which are processed by an AI Application (e.g. templates) used for authentication within the AI Mobile Device, SHALL NOT be transferred off the device.	
TS47_4.2.1_REQ_007	Users' Biometric Data (such as facial data, fingerprint data, etc.) SHALL be encrypted when at rest on the device. Encryption/decryption of this data SHALL be performed in a Secured Environment.	
TS47_4.2.1_REQ_007.1	Biometric Data SHALL also be stored in the Secured Environment.	
TS47_4.2.1_REQ_012	The temporary Voiceprint Data SHALL NOT remain in the memory after processing.	

_____(Company Representative Signature or Company Stamp)

_____(Company Representative Print Name)

_____(Company Representative Job Title)

_____(Date)

Annex D Testing Methods

D.1 Hardware performance testing either Test Method 1 or 2 can be used

D.1.1 Testing Method 1

Test Model preparation

1. Take VGG16_notop as the Reference Model.
2. Use the Model Conversion tool provided by the chipset vendor to convert the Reference Model to an int8 or/and a float16 model that can be run on the DUT, take this converted model as Model_t.
3. Validate Model_t as described in Annex E. For Model_t the validation confidence threshold is [95%].

Test Scripts preparation

Scripts to pre-process the test dataset, run the test model and measure TOPS.

Test Dataset

1000 images of size 224*224*3.

D.1.2 Testing Method 2

1. For hardware performance testing, determine the number of operations required to complete the task for the selected reference model.
2. Validate converted Test Model as described in Annex E.
3. The datasets and models shown in the following table can be used. These models and datasets can be obtained from public sources (Note).

Task	Image classification	Object detection	Image segmentation	Natural Language Processing
Data	ImageNet	COCO	ADE20K	SQuAD v1.1
Model	MobileNetEdge	MobileDET_SSD	Deeplabv3+ - MobileNetv2	MobileBERT
Scenario	Single Stream / Offline	Single Stream	Single Stream	Single Stream

NOTE: An example public source for AI test application/dataset is mlcommons.org.

Annex E Test Model Validation (Normative)

E.1 Motivation

When applying the compressed AI model to the TOPS and TOPS/w test, several concerns may arise. Firstly, the test AI models cannot be unified across DUTs since most vendors have their own methods and tools for model compression, causing models to have different formats, parameters or even structures. Secondly, the hardware design of the DUTs varies, which may lead to different processing of hardware computation.

Unfairness would thus be introduced, especially when there exist test models that are over-compressed for acceleration, or when a considerable amount of MAC operations is ignored during DUT processing. In these cases, the model would have significant information loss and its output could be too obscure for further use.

The purpose of Test Model Validation is to avoid considering the above model as valid, so as to maintain the fairness as much as possible.

E.2 Mechanism

The main idea of the validation process is to conduct a one-by-one examination on the test model output, making sure the information and utility can be retained to a certain extent after the compression and hardware processing. Comparison is made between the output of the test model running on DUT and the output of the original model (i.e., the uncompressed model) running on a trusted third-party device. If the outputs are similar, it can be considered that the test model is close to the original model, and the test model can be validated.

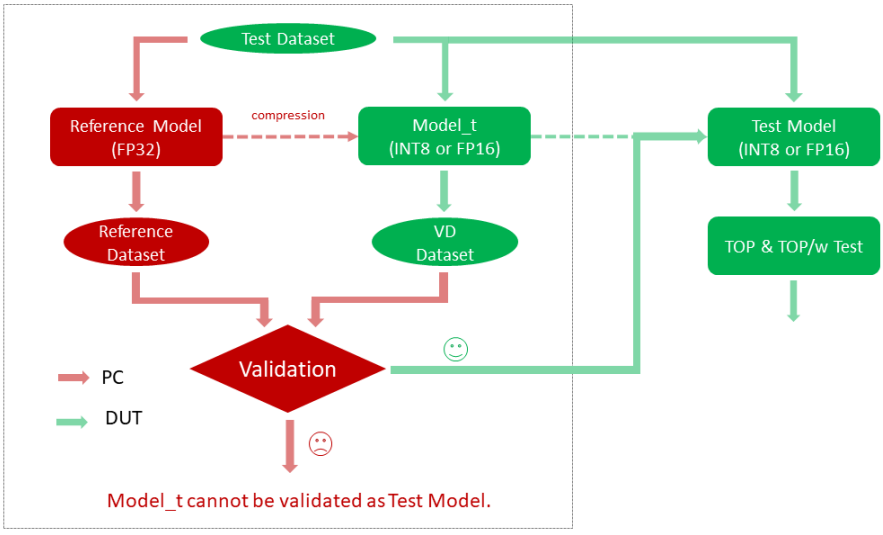

Figure E-1.. A flowchart of Test Model Validation.

As described in Section 3, the input Test Dataset can be denoted as $I = \{I_1, I_2, \dots, I_N\}$, where $N = 1000$. Consider the PC model function $F(\cdot)$ and the DUT model function $G(\cdot)$, Reference Dataset can be denoted as $R = \{R_1, R_2, \dots, R_N \mid R_n = F(I_n),\ 1 \le n \le N\}$ and VD Dataset as $V = \{V_1, V_2, \dots, V_N \mid V_n = G(I_n),\ 1 \le n \le N\}$.

Denoted the difference function as $diff(\cdot,\cdot)$, the difference matrix $DiffMat$ can be formed. The element at row n and column m is represented as

$$DiffMat[m,n] = diff(R_m, V_n), 1 \le m \le N, 1 \le n \le N,.$$

	V1	V2	...	VN
R1	diff(R1, V1)	diff(R1, V2)	...	diff(R1, VN)
R2	diff(R2, V1)	diff(R2, V2)	...	diff(R2, VN)
⋮	⋮	⋮	⋱	⋮
RN	diff(RN, V1)	diff(RN, V2)	...	diff(RN, VN)

Table E-1. An illustration of $DiffMat$ and its elements.

After $DiffMat$ is achieved, the diagonal elements $\{DiffMat[n,n] \mid 1 \le n \le N\}$ are examined. Since V_n and R_n share the same input data, theoretically they will enjoy the highest similarity. Thus, an early stopping examination is performed first, checking whether each diagonal element is the minimum of its row. Ideally, it should satisfy

$$\underset{1\le k\le N}{\operatorname{argmin}} DiffMat[n,k] = n, \qquad \forall n \epsilon [1,N].$$

But the compressed model may compress the information a lot while still preserve the class feature, directing V_n's most similar R_k to another same-class data whose index $k \neq n$, such as the case when $\underset{1\le k\le N}{\operatorname{argmin}} DiffMat[1,k] = 2$ and both V_1 and R_2 belong to the same class. So an empirical threshold is set in the test. If the proportion of minimum diagonal elements to N passes this threshold, the test continues.

Then classification is implemented to sort out the elements with strong similarity. A threshold T is set to classify all the $DiffMat$ elements. For $\forall m \epsilon [1,N], \forall n \epsilon [1,N]$,

$$\begin{cases} DiffMat[m,n]\ is\ a\ Positive\ instance, & if\ DiffMat[m,n] \le T, \\ DiffMat[m,n]\ is\ a\ Negative\ instance, & if\ DiffMat[m,n] > T. \end{cases}$$

In the final step, the F1-score of the classification is calculated. Taking all the Positive diagonal elements as $True\ Positive$ instances, the F1-score is achieved by

$$F1 = \frac{2PR}{P+R}$$

where the precision rate P and the recall rate P are computed as

$$P = \frac{True\ Positive}{True\ Positive + False\ Positive} = \frac{the\ number\ of\ True\ Positive\ instances}{the\ number\ of\ Positive\ instances}$$

$$R = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{the\ number\ of\ True\ Positive\ instances}{the\ number\ of\ diagonal\ elements}$$

Then it can be considered that under the confidence of $F1$, the information and utility preserved in the test model output are similar to the information and utility preserved in the original model's output. If the F1-score exceeds certain threshold, it can be considered that the test model and the original model have a strong similarity.

E.3 Test Setting

Specifically, Euclidean distance (i.e., l2) is recommended as the difference function, according to its better classification performance shown in Figure E-2. Structural similarity (SSIM) can be used as a difference function either.

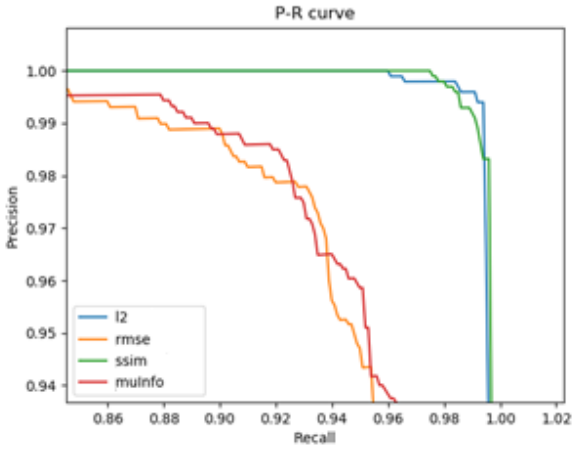

Figure E-2. The Precision-Recall curves of four candidate difference functions, where rmse denotes root mean square error and muInfo denotes mutual information. Euclidean distance and SSIM both have greater AUC (Area Under Curve) than the other two candidates.

Based on the Test Dataset in Section 3, the empirical threshold in early stopping examination is set as [99%], so that the proportion of minimum diagonal elements to N should be greater than [99%].

For the classification threshold, $T = 778.2754$ is determined by the Precision-Recall curve of a specific DUT output dataset.

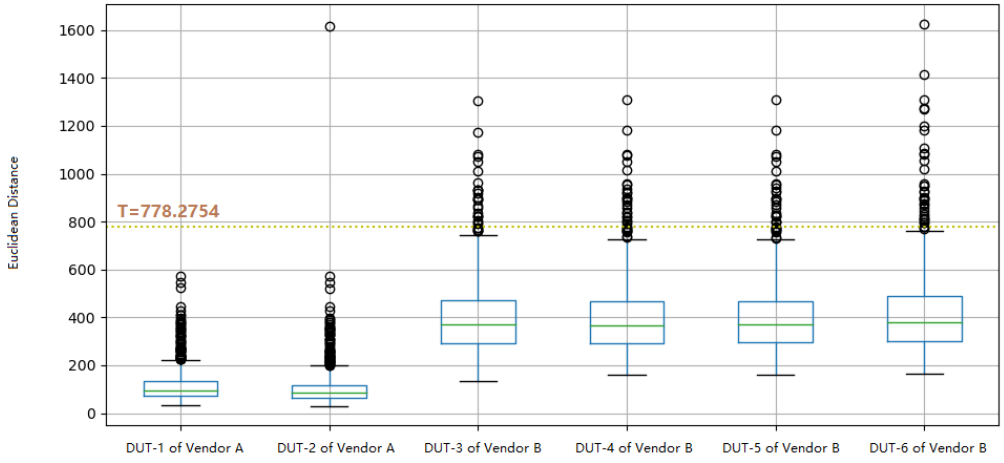

Figure E-3. The box plot of the $DiffMat$ diagonal elements of different DUTs. The majorities are below $T = 778.2754$ and all the classifications achieve F1-scores of more than 95%.
A generalised procedure is described below:

1. Number the data in Test Dataset from 1 to N and each test data item is denoted as I(n), where n ∈ [1, N].
2. Run Reference Model with Test Dataset on a PC. For each input data I(n), the corresponding output is denoted as R(n) and R(n) is in tensor form.
3. Convert Reference Model to DUT specific Test Model.
4. Run the Test Model with Test Dataset on DUT. Convert each output data item into a tensor, with the shape identical to the shape of R(n). For each input I(n) the corresponding output in tensor form is denoted V(n).
5. Calculate the difference between a given V(n) and each R(m), m ∈ [1, N]. Note: Euclidean distance is recommended as the difference function.
6. Repeat step 5 for all values if n∈ [1, N]. A two-dimensional difference matrix, DiffMat[n,n] is obtained. Each element DiffMat[m, n] represents the difference value between Reference model output R(m) and Test Model output V(n), where n, m ∈ [1, N].
7. Check whether each diagonal element is the minimum of its own row. The proportion of minimum diagonal elements in all DiffMat[n, n] should be greater than model [99%] for the model to be valid, otherwise the model is not valid and validation process stopped.
8. Classify the DiffMat elements into two classes based on their difference values. Label the elements with strong similarity as Positive (i.e., in Figure D-3, $DiffMat[m,n] \leq$ T) and the others as Negative ((i.e., in Figure D-3, $DiffMat[m,n] >$ T)).
9. Take the Positive diagonal elements as True Positive instances and calculate the F1-score of the classification. Denote the result as Fc%. The information and utility preserved in Test Model's output are similar to those preserved in Reference Model's output, under the confidence of Fc%.

If Fc% is not lower than reference model specific threshold, it is considered that DUT specific Test Model and Reference Model have a strong similarity.

Annex F Document Management

F.1 Document History

Version	Date	Brief Description of Change	Approval Authority	Editor / Company
1.0	July 2023	First version of the document	TSG#52 ISAG#	Di Zhang China Telecom

F.2 Other Information

Type	Description
Document Owner	Terminal Steering Group AI
Editor / Company	Di Zhang China Telecom

It is our intention to provide a quality product for your use. If you find any errors or omissions, please contact us with your comments. You may notify us at prd@gsma.com

Your comments or suggestions & questions are always welcome.