| Change Request Form |
| --- |
|  | TS.47 CR1001 AI Mobile Device Requirements Specification |
| Document Summary  |
| Official Document Number, Document Title and Version Number  | TS.47 AI Mobile Device Requirements Specification v0.1 (Current) |
| Official Document Type | Non-binding Permanent Reference Document |
| Change Request Security Classification | Non-confidential |
| Is this a new document or a Major or Minor Change? | New Document |
| Will this Change Request result in a Major or Minor version update? | Major Version |
| This document is for | [This document is for] |
| Input Editor and Organisation | Paul Gosden (GSMA) |
| Additional Contributors | Kay Fritz (Vodafone GmbH) |
| Issuing Group/Project | TSG |
| Approving Group/Project | TG |
| Change Request Creation Date | 12/08/2019 |
| What are the reasons for and benefits of creating this new document or Change Request? | This document defines the requirements for the implementation of AI in Mobile Devices |

© GSMA © 2019. The GSM Association (“Association”) makes no representation, warranty or undertaking (express or implied) with respect to and does not accept any responsibility for, and disclaims liability for the accuracy or completeness or timeliness of the information contained in this document. The information contained in this document may be subject to change without prior notice. This document has been classified according to the GSMA [Document Confidentiality Policy](https://infocentre-qa.concentra.co.uk/_layouts/Infocentre/InfocentreRedirect.aspx?WebId=4d56f3a7-dbc9-4a09-9a58-4aed6a0921d4&ListId=97eb8308-dd5f-422e-bbc5-8bee9731f160&ItemId=a8a9e2e4-c802-4e9c-b1f2-e85bafed771d). GSMA meetings are conducted in full compliance with the GSMA Antitrust Policy.

|  |
| --- |
| Review Log (to be completed by GSMA Support Staff) |
| Workflow Step | Document Review Comments | GSMA Support Staff Name  | Comments Date |
| Step 1: Change Request Creation (no comments required) |
| Step 2: Document Quality and/or Legal Review |
| Document Quality Team | INSERT COMMENTS HEREPlease enter details for the Quality ReviewConfirm Document Quality Team feedback Record any issues, actions and key decisions  | GSMA Support Staff Name | DD/MM/YY |
| Legal Review | INSERT COMMENTS HEREPlease enter details for the Legal Review Confirm Legal feedback Record any issues, actions and key decisions | GSMA Support Staff Name | DD/MM/YY |
| Step 3: Formal Review |
| Group(s)/Project(s) Review(s) Comments and Feedback | INSERT COMMENTS HEREPlease enter details for the Group(s)/Project(s) Review(s)Record any issues, actions and key decisionsConfirm outcome of Formal Review  | GSMA Support Staff Name | DD/MM/YY |
| Step 4: Formal Approval(s) |
| Group(s)/Project(s) Approval(s) Comments and Feedback | INSERT COMMENTS HEREPlease enter details for the Group(s)/Project(s) Approval(s)Record any issues, actions and key decisionsConfirm outcome of Formal Approval | GSMA Support Staff Name | DD/MM/YY |

# Introduction

## Purpose

This specification allows the mobile industry to design, develop, and test a new type of mobile terminal called an Artificial Intelligence (AI) Mobile Device.

This specification defines the normative baseline AI Mobile Device covering use-cases, applications, requirements and technology to accelerate the deployment of AI technology across the industry for Mobile Network Operators, devices and component manufacturers.

This specification contains normative and informative sections. Unless otherwise specified, all sections are normative.

The explanation and background information for this specification is available in the GSMA AI Mobile Device Guidelines Study Report 2018 [i.6].

## Scope

The scope of this specification is to define AI Mobile Device requirements. The AI Mobile Device in this version specifically refers to an AI mobile phone and tablet. Other types of mobile devices like IoT and wearable may be considered in future release.

##  Definition of Terms

| Term  | Description |
| --- | --- |
| Deep Learning | Deep learning is an approach to creating rich hierarchical representations through the efficient training of architectures with arbitrarily many layers. Deep learning uses multi-layered networks of simple computing units (or “neurons”). In these neural networks each unit combines a set of input values to produce an output value, which in turn is passed on to other neurons downstream. Neural networks in Deep learning are composed of several hidden layers. [Ref: ISO/IEC 23053, 3.x] |
| Deep Neural Network (DNN) | A Deep Neural Network (DNN) is created using the Deep Learning techniques defined above. |
| Facial Photo Enhancement | An application that can do one or more of the following: remove spots, reduce wrinkles, reshape facial features (such as lips, nose, cheeks, ears etc.), remove dark circles, and alter skin tone and other common imperfections when taking selfies. |
| Native API | APIs provided by the device manufacturer for access to AI hardware (e.g., NPU, CPU, GPU and DSP). |
| Native Application | An application that is pre-installed by the device manufacturer. |
| OPS | Operations Per SecondOperations only refers to multiply-accumulate (MAC) operations, not including input, output and other operations, and typically 1 MAC operation = 2 Deep Learning operations; The number of MACs needed to compute an inference on a single image is a common metric to measure the efficiency of the model. The widths of the integer matrix multiplication vary by architecture, dedicated hardware and supported topologies. Any claimed TOPS number depends on several assumptions such as frequency, number of MACs and various other hardware specifications. |
| OPS/w | OPS per watt extend that measurement to describe performance efficiency. |
| Software Framework | A software framework is a universal, reusable software environment that provides particular functionality as part of a larger software platform to facilitate development of software applications, products and solutions. Software frameworks may include support programs, compilers, code libraries, tool sets, and application programming interfaces (APIs) that bring together all the different components to enable development of a project or system. |
| TensorFlow | TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications. |
| TensorFlow Lite | TensorFlow Lite is an open source deep learning framework for on-device inference. (<https://tensorflow.org/>) |
| Third-party Applications | An application installed by the user. |

## Abbreviations

| Term  | Description |
| --- | --- |
| AI | Artificial Intelligence |
| API | Application Programming Interface |
| AR | Augmented Reality |
| ASR | Automatic Speech Recognition |
| Caffe | Convolutional Architecture for Fast Feature Embedding |
| Caffe2 | Caffe2 is a deep learning framework that provides an easy and straightforward way for experimentation with deep learning by using community contributions of new models and algorithms. Users bring their creations to scale using the power of GPUs in the cloud or to the masses on mobile with Caffe2’s cross-platform libraries. (https://caffe2.ai/docs/caffe-migration.html). |
| CPU | Central Processing Unit |
| DNN | Deep Neural Network |
| DSP | Digital Signal Processing |
| FAR | False Acceptance Rate  |
| FPE | Facial Photo Enhancement |
| FRR | False Rejection Rate |
| GPU | Graphics Processing Unit |
| GSMA | Global System for Mobile Communications, originally Group Special Mobile Association |
| MAC | Multiply-accumulate |
| MEC | Mobile Edge Computing |
| NLP | Natural Language Processing |
| NPU | Neural Processing Unit |
| SAR | Spoof Acceptance Rate  |
| SDK | Software Development Kit |
| SE | Secure Element |
| TAF | Telecommunication Terminal Industry Forum Association |
| TAR | True Acceptance Rate |
| TEE | Trusted Execution Environment |
| TOPS | Tera Operations Per Second |
| TTS | Text-To-Speech |
| VGG | Visual Geometry Group (Department of Engineering Science, University of Oxford) |

## References

Requirements SHALL be based on the exact versions as indicated below. However, if the manufacturers use a later release and/or version this SHALL be indicated. The GSMA will take efforts to continually align with other SDOs for timely information about release plans.

### Normative references

| Ref | Doc Number | Title |
| --- | --- | --- |
| 1.
 | 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>  |
|  | ISO\_IEC\_29100 | Information technology — Security techniques — Privacy frameworkAvailable at <https://standards.iso.org/ittf/PubliclyAvailableStandards/index.html> or <https://www.iso.org/committee/45306.html> |
|  | TAF-WG4-AS0026-V1.0.0 2018  | * TEE-based face recognition security evaluation method for mobile device

Available at <http://www.taf.net.cn/Association_standard_detail.aspx?Id=5b2b8d50-7ce7-47ad-b99c-d8e1b1cf7cee> |
|  | ETSI GS MEC  | Series standards, available at <https://www.etsi.org/technologies/multi-access-edge-computing> |
|  | Regulation (EU) 2016/679 | General Data Protection Regulation Available at https://gdpr-info.eu/ |
|  | GPD\_SPE\_009 | TEE System ArchitectureAvailable at<https://globalplatform.org/specs-library/> |
|  | GB/T 35273-2017 | Information security techniques - Personal information security specificationAvailable at https://www.tc260.org.cn/front/bzcx/yfgbcx.html |
|  | ISO/IEC 29101 | 《Information technology — Security techniques — Privacy architecture framework》Available at<https://www.iso.org/standard/75293.html> |
|  | PUBLIC LAW 106–102—NOV. 12, 1999 | Gramm-Leach-Bliley Act (USA)Available athttps://www.congress.gov/106/plaws/publ102/PLAW-106publ102.pdf |
|  | GB/T36464.4—2018 | Information technology—Intelligent speech interaction system—Part4:Mobile terminalAvailable at<http://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=7659A1A0BF2EE19723B46BC159057572> |

### Informative references

| Ref | Title |
| --- | --- |
| [1] | Wang, Shiqiang, et al. "When edge meets learning: Adaptive control for resource-constrained distributed machine learning." IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018. |
| [2] | Huang, Kaibin, et al. "Communication, Computing, and Learning on the Edge." 2018 IEEE International Conference on Communication Systems (ICCS). IEEE, 2019. |
| [3] | Mao, Yuyi, et al. "A survey on mobile edge computing: The communication perspective." IEEE Communications Surveys & Tutorials 19.4 (2017): 2322-2358. |
| [4] | Nazer, Bobak, and Michael Gastpar. "Computation over multiple-access channels." IEEE Transactions on information theory 53.10 (2007): 3498-3516. |
| [5] | Zhu, Guangxu, and Kaibin Huang. "MIMO over-the-air computation for high-mobility multi-modal sensing." IEEE Internet of Things Journal (2018). |
| [6] | Study Report of AI Mobile Device Guidelines <https://infocentre2.gsma.com/gp/wg/TS/WorkingDocuments/TSG33_035%20TSG%20Study%20Report%20of%20AI%20Mobile%20Device%20Guidelines%20v2.0.docx> |

## 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 [1].

The word *SHALL* indicates mandatory requirements strictly to be followed in order to conform to the standard and from which no deviation is permitted (*SHALL* equals *is required to*).

Note that the use of the word *MUST* is deprecated and shall not be used when stating mandatory requirements; *MUST* is used only to describe unavoidable situations. The use of the word *WILL* is deprecated and shall not be used when stating mandatory requirements; *WILL* is only used in statements of fact.

The word *SHOULD* indicates that among several possibilities, one is recommended as particularly suitable without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required (*SHOULD* equals *is recommended that*).

The word *MAY* is used to indicate a course of action permissible within the limits of the standard (*MAY* equals *is permitted to*).

# AI Mobile Device Definition

An AI mobile device refers to a mobile device that has the following characteristics:

1. On-device computational resources to enable AI deep learning and other AI algorithms based on either dedicated AI hardware or general hardware to support deep learning AI applications.
2. On-device software framework to support the updating of AI deep learning neural networks
3. On-device AI software to perform inferencing using deep neural network models

# The Requirements of AI Mobile Device

## Hardware requirements

AI mobile device hardware is required to support AI software applications efficiently.

Example hardware performance measurements can be found in the Table below TS.47\_3.1\_REQ\_001 to TS.47\_3.1\_REQ\_004 using the modified VGG network.

|  |  |
| --- | --- |
| TS47\_3.1\_REQ\_001 | An AI mobile device SHALL have a minimum of [1] int8 TOPS. |
| TS47\_3.1\_REQ\_002 | An AI mobile device SHALL have a minimum of [0.5] float16 TOPS |
| TS47\_3.1\_REQ\_003 | An AI mobile device SHALL have a minimum of [0.5] int8 TOPS/Watt. |
| TS47\_3.1\_REQ\_004 | An AI mobile device SHALL have a minimum of [0.3] float16 TOPS/Watt. |

## Software requirements

AI mobile device software requirements:

|  |  |
| --- | --- |
| TS47\_3.2\_REQ\_001 | An AI mobile device SHALL support on-device model updates of an existing deep learning network. |
| TS47\_3.2\_REQ\_002 | An AI mobile device SHALL support native APIs to expose the AI hardware functions. |
| 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. |
| 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, \*.jason and \*.onnx. |
| TS47\_3.2\_REQ\_005 | An AI mobile device SHOULD provide an SDK to support definition of new customized Deep Learning operators. |

For the existing SDKs and APIs refer to Annex A.1.

## Deep Learning Application Requirements

Deep Learning applications include but not limited to biometric functions, image processing, speech, augmented reality (AR) and system optimization categories.

### Biometric Performance Requirements

|  |  |
| --- | --- |
| TS47\_3.3.1\_REQ\_001 | An AI mobile device SHALL implement and certify one or more of the biometric systems defined by TS47\_3.3.1\_REQ\_001.1, TS47\_3.3.1\_REQ\_001.2 and TS47\_3.3.1\_REQ\_001.3. |
| TS47\_3.3.1\_REQ\_001.1 | An AI mobile device SHOULD support a fingerprint biometric system. |
| TS47\_3.3.1\_REQ\_001.2 | An AI mobile device SHOULD support a 2D facial biometric system. |
| TS47\_3.3.1\_REQ\_001.3 | An AI mobile device SHOULD support a 3D facial biometric system. |
| TS47\_3.3.1\_REQ\_002 | 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 ProgramInternet Finance Authentication Alliance (IFAA) biometric Certification Program |
| TS47\_3.3.1\_REQ\_003 | An AI mobile device supporting 2D facial biometric system SHALL support the biometric KPI requirement TS47\_3.3.1\_REQ\_003.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization. |
| TS47\_3.3.1\_REQ\_003.1 | 2D Facial FAR <= [0.002]% and FRR <= [3]% simultaneously |
| TS47\_3.3.1\_REQ\_004 | An AI mobile device supporting 3D facial biometric system SHALL support the biometric KPI requirement TS47\_3.3.1\_REQ\_004.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization. |
| TS47\_3.3.1\_REQ\_004.1 | 3D Facial FAR <= [0.001]% and FRR <= [3]% simultaneously |
| TS47\_3.3.1\_REQ\_005 | An AI mobile device supporting fingerprint biometric system SHALL support the biometric KPI requirement TS47\_3.3.1\_REQ\_005.1 for each of the use cases: Device Unlock, Application Login and Payment Authorization. |
| TS47\_3.3.1\_REQ\_005.1 | Fingerprint FAR <= [0.002]% and FRR <= [3]% simultaneously |

### On-Device Image Processing Requirements

|  |  |
| --- | --- |
| TS47\_3.3.2\_REQ\_001 | An AI mobile device SHALL have on-device computer vision capabilities which can be directly used by native and third-party applications through an OEM specific API. |
| TS47\_3.3.2\_REQ\_002 | An AI mobile device SHALL have optical character recognition (OCR) capability on the device. |
| TS47\_3.3.2\_REQ\_003 | An AI mobile device SHALL have image detection, image classification and image segmentation capabilities on the device. |
| TS47\_3.3.2\_REQ\_004 | An AI mobile device SHALL have face detection and face clustering capabilities within a group of photos on the device. |
| TS47\_3.3.2\_REQ\_005 | An AI mobile device SHALL have video super-resolution capabilities on the device. |
| TS47\_3.3.2\_REQ\_006 | An AI mobile device SHALL have video classification capabilities on the device. |

#### On-Device Image Processing Applications

|  |  |
| --- | --- |
| TS47\_3.3.2.1\_REQ\_001 | The AI Mobile Device SHALL support all of the following applications:Photo scene detection and recognition1. Identification of one or more objects in different scenes such as portraits, landscapes, foods, night scenes and texts, etc.
2. Scene detection capabilities to optimize camera settings for image capture based on scene content.

Text detection and recognition of downloaded languages:1. Different languages.
2. In natural scenes, such as the text on billboards, menus, vehicle license plate, and product descriptions.
3. Of business cards, ID cards, passports, driver licenses, and credit cards.
 |
| TS47\_3.3.2.1\_REQ\_002 | The AI Mobile Device SHOULD support, automatic language detection |
| TS47\_3.3.2.1\_REQ\_003 | The AI Mobile Device SHOULD provide personalized FPE for users based on gender, age, and skin tone. |
| TS47\_3.3.2.1\_REQ\_004 | The AI Mobile Device SHOULD support FPE of multiple people in a single photo. |
| TS47\_3.3.2.1\_REQ\_005 | The AI Mobile Device SHOULD support user adjustment of the FPE level from no enhancement to the max FPE. |
| TS47\_3.3.2.1\_REQ\_006 | The AI Mobile Device SHOULD support automatic classification of photos in an album by different categories. |

### Speech

|  |  |
| --- | --- |
| TS47\_3.3.3\_REQ\_001 | The AI mobile device SHALL have speech ability, including but not limited to voice assistant. |

#### Voice assistant

|  |  |
| --- | --- |
| TS47\_3.3.3.1\_REQ\_001 | AI mobile device SHALL support voice assistant function.  |
| TS47\_3.3.3.1\_REQ\_002 | The AI mobile device SHALL provide Automatic speech recognition (ASR) capabilities. |
| TS47\_3.3.3.1\_REQ\_003 | The AI mobile device SHALL provide Natural Language Understanding (NLU) capabilities. |
| TS47\_3.3.3.1\_REQ\_004 | The AI mobile device SHALL provide Synthesized Voice (TTS)) capabilities. |
| TS47\_3.3.3.1\_REQ\_005 | The AI mobile device SHALL support voice trigger |
| TS47\_3.3.3.1\_REQ\_006 | It SHOULD support voiceprint recognition for preventing people other than the device’s owner from triggering voice assistant. |
| TS47\_3.3.3.1\_REQ\_006.1 | In a quiet environment, the follow SHALL be required:The true acceptance rate (TAR) >=[90]%, and the false acceptance rate (FAR) of voiceprint recognition <= [20]%. |
| TS47\_3.3.3.1\_REQ\_006.2 | In a noise environment, the follow SHALL be required:TAR >=[80]%, and FAR of voiceprint recognition <= [20]%. |
| TS47\_3.3.3.1\_REQ\_007 | 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).  |
| TS47\_3.3.3.1\_REQ\_008 | The AI mobile device SHOULD have access to different categories of applications and invoke these applications’ services and functions via voice assistant. |
| TS47\_3.3.3.1\_REQ\_009 | The AI mobile device SHALL support information search by on-device voice assistant. |
| TS47\_3.3.3.1\_REQ\_010 | The AI mobile device SHOULD support interaction with smart devices (e.g. home appliances) via voice assistant. |

### Augmented Reality (AR)

|  |  |
| --- | --- |
| TS47\_3.3.4\_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
 |
| TS47\_3.3.4\_REQ\_002 | The AI Mobile device SHOULD support the following applications:1. AR Emoji
	1. Creating customized AR-based Emoji.
	2. Tracking user’s facial movement and expression and render these on the AR-based Emoji.
2. AR video
	1. Compositing real objects with virtual objects and/or virtual background
	2. Minimum [30] fps frame rate
	3. AR shadow effect and occlusion handling.
	4. AR enhanced information text labels should not deviate or disappear from the actual target scene when the AI mobile device moves.
 |

### System Optimization

|  |  |
| --- | --- |
| TS47\_3.3.5\_REQ\_001 | The AI Mobile Device SHOULD support dynamic system resource allocation and optimization based on feedback provided by on-device sensors measuring environmental conditions combined with continuous learning of user habits and behaviours:1. Dynamic application management (e.g. pre-loading, closing, put to sleep, control network access) based on user’s habits (e.g. usage duration, frequency).2. Dynamic application management based on abnormal behaviour detection (e.g. increased memory usage, abnormal power consumption, self-starting in the background) 3. Dynamic system resource management based on continuous learning of system performance (e.g. memory and storage defragmentation, off-line storage during off-peak periods).4. Dynamic system resource allocation for high performance applications (e.g., gaming and video). |

# AI Agent (informative)

Another crowning achievement of deep learning is its extension to the domain of reinforcement learning. In the context of reinforcement learning, an autonomous agent SHALL learn to perform a task by trial and error, with minimal guidance from the user.

Examples of AI agent capabilities but not limited to:

1. An agent is responsible for the decision-making of AI computation offloading, and implements an MEC-first strategy, i.e. abstracting the computation offloading decision function from specific application, and make it become a functional entity on AI mobile device.

On-device deep reinforcement learning will enable device to perceive the environment and react autonomously. Supporting more and more autonomous applications will be the trend, which will make AI mobile device significantly different from smartphone today.

AI agents are software entities which can carry out some actions on behalf of clients with some degree of autonomy.

In general, agents possess five common properties which are autonomy (some level of self-control), adaptively (the ability to learn and improve performance with experience), reactivity (the ability to perceive the environment and to respond in a timely fashion to changes that occur), proactivity (the ability not only to act simply in response to their environment but also to exhibit goal-directed behaviour by taking the initiative) and sociability (the ability to interact, communicate and work with other agents).

AI agent will dramatically change the landscape of the mobile device. It can act as the “brain” of the mobile device, to control the behaviour and system performance of the device. It can act as the new “entrance of services”, recommend services (applications) to the end user based on the context. It can interact with other agents; the communication between AI agents can achieve the cross-device inference.

In the future, the AI agent will become an important feature for defining an AI mobile device.

## Privacy and security requirements for AI agent (informative)

The user shall be informed about how the AI agent may affect them.

The user shall be able to lodge a complaint against processing by AI agent, or to consent to it.

The decisions and recommendations made by the AI agent shall be understandable by a user.

The decisions and recommendations made by the AI agent shall be explained in a way that the user is able to understand the result.

The user shall be informed how to oppose and override the decision made by AI agent.

# Network Requirements to Support AI Mobile Devices (informative)

Computation on AI mobile devices may be improved by offloading to MEC or Cloud to reduce latency and mobile power consumption.  The ubiquitous AI mobile device will make AI computation a very important task for the network to bear, which will ultimately drive the network to change.

1. Cloud computing centres mayhave the ability to provide AI as a service.
2. MEC mayhave the ability to provide AI as a service, which is equivalent to location service, bandwidth management service and radio network information service, and provide unified open APIs [4].
3. Networks may gradually evolve from a communication platform to a platform that supports both communication and computation, in order to better support edge learning.

# Privacy and Security Requirements

## Privacy Requirements

Applicable law(s) as related to privacy should be complied with in connection with AI on mobile devices.

## Security Requirements

Applicable law(s) as related to security should be complied with in connection with AI on mobile devices.

|  |  |
| --- | --- |
| TS47\_6.2\_REQ\_001 | The use of AI on mobile devices SHALL adhere to the security requirements from the GSMA Mobile Privacy Principles that personal information must be protected, and the AI Mobile Device SHALL use reasonable safeguards appropriate to the privacy, sensitivity, confidentiality and integrity of the information. |

### Security for AI applications

|  |  |
| --- | --- |
| TS47\_6.2.1\_REQ\_001 | The security and the robustness of the AI models used by the AI mobile devices SHOULD be guaranteed with appropriate safeguards to protect and prevent Confidentiality, Integrity and Replay attacks. |
| TS47\_6.2.1\_REQ\_002 | Defence techniques SHOULD take into account for protecting models' training from confidentiality and integrity attacks. For example, in evasion attacks, data can be manipulated to mislead AI models. |

For AI applications with high security requirements, defence techniques are recommended to use on AI models as follows:

Defence techniques (e.g. network distillation, adversarial training, adversarial sample detection, etc.) are recommended to use on AI models to prevent them from evasion attacks.

Defence techniques (e.g. training data filtering, regression analysis, ensemble analysis, etc.) are recommended to use on AI models to prevent them from poisoning attacks.

Defence techniques (e.g. high-intensity encryption algorithm, input preprocessing, model pruning, etc.) are recommended to be used on AI models to prevent them from backdoor attacks.

An AI agent shall be protected from external threats through a platform of system vulnerability protection with a service which protects the platform and legitimate agents from insecure internal processes.  A secure platform will provide basic security measures of authentication, authorization, availability, confidentiality and integrity.

1. Biometric Authentication

|  |  |
| --- | --- |
| TS47\_6.2.1\_REQ\_003 | Users' biometric data (such as facial data, fingerprint data, etc.) SHALL be encrypted. Encryption/decryption of the data SHALL be done in secure unit, and key materials SHALL also be stored in secure unit (SE) [6]. |
| TS47\_6.2.1\_REQ\_004 | Biometric algorithms (such as face recognition algorithms, fingerprint algorithms, etc.) SHALL run in a private and secure execution environment such as trusted execution environment (TEE) [6]. |
| TS47\_6.2.1\_REQ\_005 | If users' biometric data is replaced, the previous biometric data before the replacement SHALL be deleted completely and permanently, and shall not be recovered by data rollback. |
| TS47\_6.2.1\_REQ\_008 | The biometric data SHALL also be wiped and made unrecoverable by a device factory reset. |

1. Speech

|  |  |
| --- | --- |
| TS47\_6.2.1\_REQ\_006 | Voiceprint data SHOULD be stored on the device with encryption. |
| TS47\_6.2.1\_REQ\_009 | The temporary voiceprint data shall not remain in the memory after processing. |
| TS47\_6.2.1\_REQ\_010 | When the voiceprint data is permanently and completely deleted, it SHALL not be recovered by data rollback. |
| TS47\_6.2.1\_REQ\_011 | The voiceprint data SHALL also be wiped and made unrecoverable by a device factory reset. |

1. Augmented Reality

|  |  |
| --- | --- |
| TS47\_6.2.1\_REQ\_007 | Appropriate safeguards SHOULD be used to protect AR applications from malicious application attacks, such as spoofing user with information about the real and/or virtual world, sensory overload attacks, hijacking users' clicks, etc. |

1. Informative
	1. SDK&API

Currently, each chipset vendor has its own set of APIs, which lead to fragmented ecosystem. To standardize and unify application APIs are very necessary and highly recommended.

* + 1. The Android Neural Networks API (NNAPI)

The Android Neural Networks API (NNAPI) is an Android C API designed for running computationally intensive operations for machine learning on mobile devices. NNAPI is designed to provide a base layer of functionality for higher-level machine learning frameworks (such as TensorFlow Lite, Caffe2, or others) that build and train neural networks.

< Official website url, <https://developer.android.com/ndk/downloads>>

* + 1. The Snapdragon Neural Processing Engine (SNPE)

The Snapdragon Neural Processing Engine (SNPE) is a Qualcomm Snapdragon software accelerated runtime for the execution of deep neural networks. The Qualcomm Neural Processing SDK for artificial intelligence (AI) is designed to help developers run one or more neural network models trained in Caffe/Caffe2, ONNX, or TensorFlow on Snapdragon mobile platforms, whether that is the CPU, GPU or DSP.

Official website url, <https://developer.qualcomm.com/software/qualcomm-neural-processing-sdk>

* + 1. HiAI

HiAI is a mobile terminal–oriented artificial intelligence (AI) computing platform that constructs three layers of ecology: service capability openness, application capability openness, and chip capability openness. The three-layer open platform that integrates terminals, chips, and the cloud brings more extraordinary experience for users and developers.

Official website url, <https://developer.huawei.com/consumer/en/devservice/doc/2020301>

* + 1. NeuroPilot

NeuroPilot is MediaTek's AI ecosystem. It embraces the advantages of 'Edge AI', which means the AI processing is done on-device rather than relying on a fast internet connection and Cloud service. However, NeuroPilot doesn't have to use a dedicated AI processor. Its software can intelligently detect what compute resources are available, between CPU, GPU and APU, and automatically choose the best one.

* + 1. Core ML

Core ML is an Apple framework that allows developers to easily integrate machine learning (ML) models into apps. Core ML is available on iOS, watchOS, macOS, and tvOS. Core ML introduces a public file format (.mlmodel) for a broad set of ML methods including deep neural networks (convolutional and recurrent), tree ensembles (boosted trees, random forest, decision trees), and generalized linear models.

Official website url, <https://developer.apple.com/documentation/coreml>

* + 1. MACE

Mobile AI Compute Engine (MACE) is a deep learning inference framework optimized for mobile heterogeneous computing on Android, iOS, Linux and Windows devices. The design focuses on the following targets:

* Performance: Runtime is optimized with NEON, OpenCL and Hexagon, and Winograd algorithm is introduced to speed up convolution operations. The initialization is also optimized to be faster.
* Power consumption: Chip dependent power options like big.LITTLE scheduling, Adreno GPU hints are included as advanced APIs.
* Responsiveness: UI responsiveness guarantee is sometimes obligatory when running a model. Mechanism like automatically breaking OpenCL kernel into small units is introduced to allow better preemption for the UI rendering task.
* Memory usage and library footprint: Graph level memory allocation optimization and buffer reuse are supported. The core library tries to keep minimum external dependencies to keep the library footprint small.
* Model protection: Model protection has been the highest priority since the beginning of the design. Various techniques are introduced like converting models to C++ code and literal obfuscations.
* Platform coverage: Good coverage of recent Qualcomm, MediaTek, Pinecone and other ARM based chips. CPU runtime supports Android, iOS and Linux.
* Rich model formats support: TensorFlow, Caffe and ONNX model formats are supported.​

Official website url, <https://github.com/XiaoMi/mace>

Document Management

* 1. Document History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Date | Brief Description of Change | Approval Authority | Editor / Company |
| 1.0 | Oct 2019 | New PRD  | TSG#37TG Oct 2019 |  |

* 1. Other Information

|  |  |
| --- | --- |
| Type | Description |
| Document Owner | Terminal Steering Group (TSG) |
| Editor / Company | Kay Fritz / Vodafone |