AI ITA Nomenclature

MDIA
Malta Digital Innovation Authority

AI-ITA Nomenclature

Digital Innovation is, by definition, a rapidly evolving sector. These guidelines are expected to be updated to keep abreast with technology, regulatory and operational developments.

Document Version: 03 October 2019

Contents

1. Definitions

"Applicant", within the context of this document, refers to an individual and/or legal organisation applying for Certification of an Innovative Technology Arrangement (ITA) with the Authority.

"Authority" refers to the Malta Digital Innovation Authority ('MDIA'), as defined by the Malta Digital Innovation Authority Act, 2018 ('MDIA Act).

"Blueprint" refers to a document that includes a description of the qualities, attributes, features, behaviours or aspects of an ITA as defined in the 'ITA Blueprint Guidelines'.

"Innovative Technology Arrangement", also referred to as 'ITA' within this document, as defined within the First Schedule of the Innovative Technology Arrangements and Services Act, 2018.

"AI-ITA" refers to Innovative Technology Arrangements that exhibit features or qualities of Artificial Intelligence as recognised by the Authority and described in the 'AI Innovative Technology Arrangements Guidelines'.

"ITAS Act" refers to the Innovative Technology Arrangements and Services Act, 2018.

"Technical Administrator" ('TA') as defined in the *Innovative Technology Arrangements and Services Act, 2018*, and in line with further guidance issued by the Authority under Chapter 3 of the Guidance Notes.

"Forensic Node" as defined by the Authority within the 'Forensic Node Guidelines".

"Ethical and Trustworthy AI Framework" refers to the *Malta Towards Trustworthy AI: Malta Ethical AI Framework* guidelines, published by the *Malta.AI* Taskforce on https://malta.ai/

2. Introduction

AI systems are defined by a wide range of characteristics and features. For the purposes of the AI-ITA Certification framework, the MDIA is considering systems that apply AI techniques in order to produce AI applications as further described in the 'AI Innovative Technology Arrangement Guidelines'.

This document is intended to provide an overview of the processes and components that make up a typical AI-ITA for the purpose of describing the AI-ITA as required in the Blueprint (refer to the 'AI-ITA Blueprint Guidelines'). The Authority recognises that Artificial Intelligence is a rapidly evolving field, and the document is meant to serve as a high-level guideline into what would typically make up an AI-ITA.

Applicants who would like further clarification on whether their solution may be classified as an AI-ITA are invited to contact the Authority.

The below is an abstract process-map for the purposes of defining the high-level components that characterise AI systems. The process-map is meant to serve as a general guide into describing the different processes for an AI-ITA.

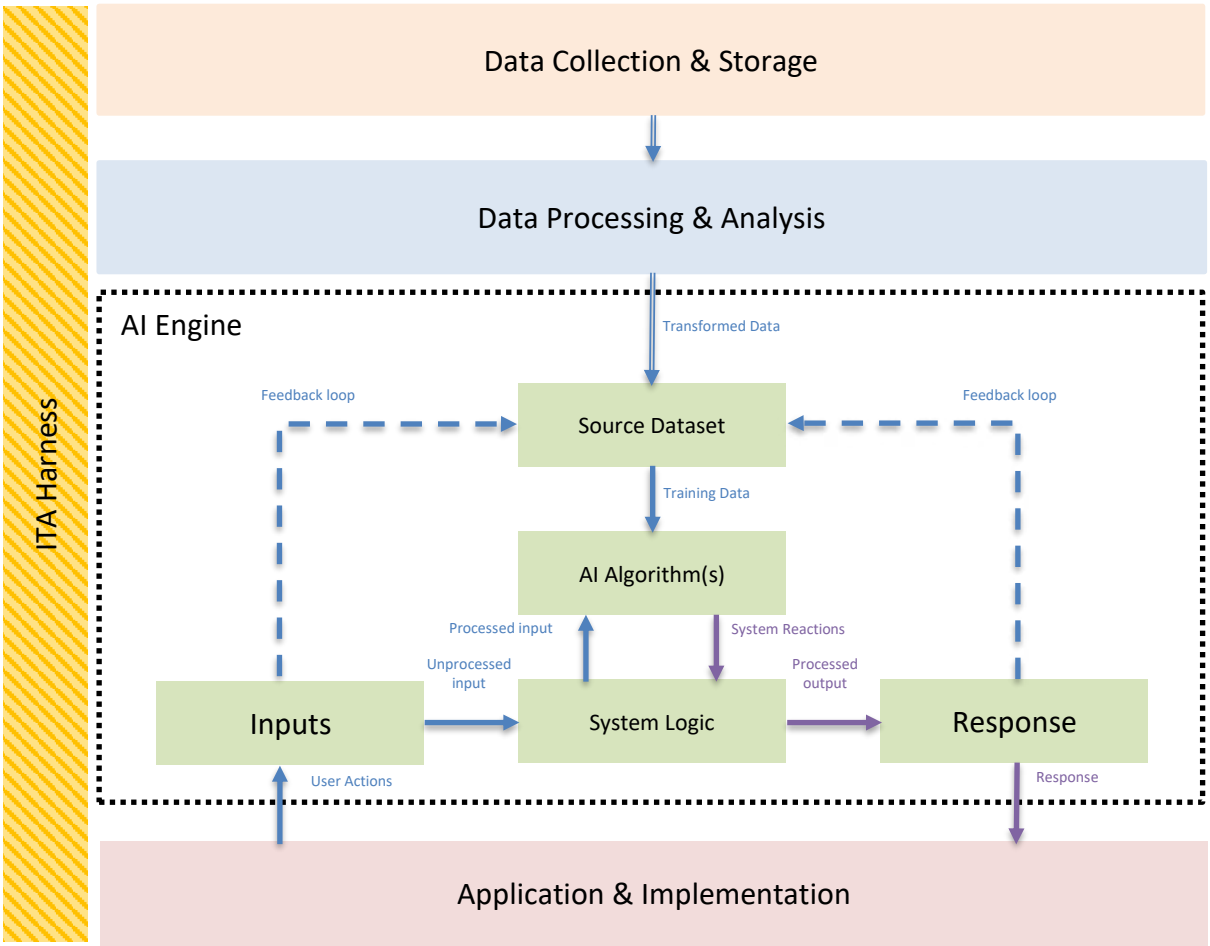

Figure 1 - AI-ITA Process-Map

3. Data Collection & Storage Process

The *Data Collection and Storage Process* defines the activity surrounding any collection, retrieval or collation of data sources that is used by the AI-ITA. Data defined within this process can be fed into the AI system for uses such as training and classification after processing and may also be used for other purposes that extend beyond the AI engine, such as user profiles to grant access to the system.

This process is very often used and relied upon by most categories of AI systems. Example uses include the basis of the training set for machine learning; data sources for multimedia to be used in Computer Vision; Corpora of text for Natural Language Processing; and Storage of human-generated knowledge upon which an Expert System will infer results.

The data may be sourced from different locations and repositories, which may be specific to an AI-ITA, from open data sources, or a combination thereof. Data may also be collected and maintained at various processes of AI-ITA, for example to store the results generated by the AI system for use by other processes and components.

This process may interact with sensitive data, including personal data, which must be handled appropriately (e.g. only made accessible to those having the necessary authorisation levels).

The ITA Harness may, wherever applicable, be used to provide an additional layer of protection against unauthorised access or modification of the data located within the boundaries of this process. Additionally, the Forensic node may also be utilised to ensure a log of access and operations carried out on the data set to enhance auditability and traceability.

The data may be stored on various platforms, such as local infrastructure, data centres and/or cloud repositories.

4. Data Processing & Analysis Process

The *Data Processing & Analysis Process* refers to any operations applied to process and derive any metrics from the underlying data source. For the purposes of the generic AI-ITA outlined in the diagram, this process may be used and relied upon by the *Data Collection & Storage Process*.

This phase may involve both automated and manual actions (or hybrids thereof), as in the case of manual classification of datasets for Machine Learning. Additionally, when the solution contains sensitive data, the process could, wherever possible, include aspects of anonymisation to allow for the personal data to be discarded and mitigate risks.

Wherever the processing and analysis applied is automated, the ITA Harness may be used to check and verify against unexpected behaviour.

5. AI Engine Process

The *AI Engine Process* represents the core of the AI-ITA system and is where the AI Algorithm(s) and surrounding logic and behaviour is located. For the purposes of the Blueprint and as illustrated in the process-map diagram (section 2. Introduction), this process must be defined at a lower level of abstraction.

This process specifies the behaviour by which the AI-ITA is initiated, how external actions triggered by users or devices interact with the AI Algorithms, and how the AI Algorithm produces and delivers the results back to the relevant process (which for the purposes of the diagram is illustrated by the *Application & Implementation Process*).

The components defined in the diagram and their purpose is outlined below.

- **Source Dataset**: is where the AI's core dataset is transferred into. In the example of a *Neural Network* this would be the training set, in the case of an *Evolutionary Algorithm* this would be the initial population.
- **AI Algorithm(s)**: is the core of the "*AI Engine* process" and is where the *Neural Network* would reside and initially be trained, or the *Evolutionary Algorithm* and related fitness functions and constraints are defined.
- **System Logic**: is where the interaction to and from the *AI Algorithm(s)* for inputs and outputs respectively and any other logic (such as switching between different algorithms) is managed and maintained.
- **Inputs**: is the component which handles the user data in a structured manner and prepares it to be passed to the System Logic module. In the event that this will be added to the data source for continuous updating of the data set which the *AI Algorithm(s)* rely upon (feedback loop), this module may also pass the data in the requisite structure to the *Source Dataset*.
- **Outputs**: is the component which takes the results and formats them for passing to the next process and/or back to the *Source Dataset* in the event of a feedback loop.

The Input and Output components may also complement the ITA Harness to carry out checks to verify that the inputs that have been passed are within the expected boundaries.

Other than the *Data* source and *AI Algorithm(s)* components which are required in all blueprints, different components altogether may be included depending on the architecture of each component.

6. Application & Implementation Process

This process defines the components outside the *AI Engine Process* that end-users would access to interface with the *AI Engine Process*.

This includes but is not limited to web applications, desktop applications, or mobile applications, and as such can be composed of a variety of technologies.

The *Application & Implementation Process* may also be composed of multiple separate processes sitting side-by-side, should there be multiple different and distinct mechanisms for the end-user to interact with the *AI Engine Process*.

7. ITA Harness

The ITA Harness component's purpose is to contain the boundaries of the operations of the AI-ITA to ensure that it behaves in a safe manner and within the expected working range. It spans across all processes as it may be invoked at various stages, including in between the processes themselves as well as embedded within one or more specific processes or components within the process.

The harness is intended to safeguard the operation of the AI-ITA by actively checking that the user actions (inputs), system reactions (outputs), and any other behaviour is within the expected boundaries of the operation of the AI-ITA and any exceptions trigger the appropriate failure modes as defined in the 'AI Innovative Technology Arrangement Guidelines'. The harness must implement checks, wherever possible, to ensure that the operation of the AI-ITA is in line with the 'Ethical and Trustworthy AI Framework', by for example ensuring that behaviour that may cause harm is automatically detected and the necessary fallback plans are automatically put in action.

Other than an overarching ITA Harness, complementary components to the harness may be applied within the implementation of every process itself. As a general example and in line with the abstract process-map, consider the *AI Engine Process* which may complement the ITA Harness by carrying out checks on:

1. The data source as it is received from the *Data Process & Analysis Process* to detect and identify possible data pollution, bias or anomalies that may result in unexpected failures once in use by the *AI Engine Process*;
2. The user actions and any other inputs to ensure they are in line within the expected range;
3. The system reactions and outputs to ensure that they are within the boundaries of expected results from the *AI Engine Process*.

Decisions taken by the ITA Harness must be logged to the Forensic node as detailed in the 'AI-ITA Forensic Node Guidelines' unless valid justification for not doing so is given, to ensure traceability of the inputs and outputs at the very minimum.

There may be instances where a harness is limited or not applicable to an AI-ITA as it may not be deemed required, feasible or desirable, in which case justification and associated risks (as well as alternate mitigation measures) must be described in detail within the Blueprint to ensure that any operating risks of the AI-ITA are well defined, monitored and contained separately.

The ITA Harness is not intended to be a replacement for human-level checks, and the responsibility still lies with the designated Technical Administrator. It is however intended to augment capabilities and aid in detection of possibly unexpected behaviour before the fact.

8. Variations

As AI-ITAs will vary significantly from one another and may utilise entirely different AI algorithms and architectures, the process-map defined above is intended to be used as a generic guideline.

Each AI-ITA may utilise different processes and components, which due to the specificity of each AI-ITA must be clearly defined and described in the Blueprint.

9. Boundaries, Dependencies and Interfaces

When describing the Innovative Technology Arrangement (ITA) it is important to define:

i. the boundaries of the ITA, i.e. the parts of the ITA that fall within a specific process or across different process modules;
ii. any dependencies which the ITA relies on and their implications; and
iii. the interfaces which each module will expose for use by other layers or users (whereby users could also be other computational systems);