Chapter 2 - Overview of Testing AI Systems

Testing of AI Systems

- The use of AI in various fields is increasing.
- The theory around how to test AI is in its nascent stage.
- Testing could be performed at the stage of training the algorithm, as in the design cycle or as part of the system testing.
- We will treat the algorithm as a black box and look at testing as part of system testing.

Testing of AI-ML Systems

Testing AI systems is a two-stage process:

- Testing the trained model (offline).
- Testing the integrated system – AI and non-AI system components (Online).

2.1 AI Testing Phases

2.1.1 Offline and Online Testing of AI Systems

The ML lifecycle, as described in 1.4 Stages of the ML Process, can be divided into two phases: offline and online. Distinct types of tests performed during these phases are:

Offline phase testing: In this phase, the trained model is tested. Various metrics are used for evaluation parameters for a trained model to verify how far it has achieved the objectives. There are different parameters for both supervised and unsupervised learning. Typically, the model is tested for functional behavior but non-functional characteristics are not tested. Since the model is deployed in an environment which is different from the training environment, doing performance testing of the model in this phase does not make much sense. Model training time is a parameter that is evaluated as a non-functional parameter. For models that will need to be retrained frequently, this can be an important parameter. Offline testing is covered in Chapter 3 - Offline Testing of AI Systems.

Another Important aspect of offline testing is whether it is possible to be able to explain the behavior of the model. There are various methods and algorithms that can be used for it. This aspect of testing is covered in Chapter 5 - Explainable AI.

Online phase testing: In this phase, the integration of the trained model with the rest of the system, including all other AI and non-AI components, is tested. Both functional and non-functional tests such as performance tests can be performed.

Since the inputs to the system can be non-textual, unstructured inputs, as well automation support for some of the tests related to the ML part, may be limited, based on the tool being used.

Online testing is covered in Chapter 4 - Online Testing of AI Systems.

2.2 AI vs. Non-AI Testing

2.2.1 Testing of AI systems vs. Traditional (non-AI) Systems

- Testing AI Systems is not deterministic. The results of the test are probabilities.
- Test oracles for AI systems are not easily available
- Data are the test cases for the AI.
- Data pre-processing and clean up constitute a vital part of AI system testing.
- The internals of learning systems are generally not easy to understand.
- The explainability of the results is difficult for ML and nearly impossible for DL systems.
- AI systems logic is generated based on the data used to train the mode.
- That logic is not available for examination, especially neural nets. This makes it difficult to understand why a particular output was produced. A correct or desired answer doesn't guarantee correct functioning.
- Testing in the offline phase is an additional step which requires specialized skills and techniques for testing the trained model.
- Testing in the online phase requires a deep understanding of how AI systems work and how to integrate these with other AI and non-AI systems. As a consequence, it calls for an increased for diverse test design techniques.
- Black-box testing methods and conventional testing are also applicable to AI based systems testing.
- Similar to non-AI systems, both functional and non-functional tests need to be executed for AI systems.
- Online phase testing can be performed as normal black-box system and system integration testing without worrying whether there are one or more AI components in the mix.

Test cases for AI Systems

- The behaviors of Learning Systems depend on the input data.
- Different data will lead to potentially different behaviors.
- For AI systems, test data = test cases.
- Typically, one test case in traditional systems can have multiple data items.

2.3 AI Quality Characteristics

2.3.1 Quality Characteristics for Evaluating AI Systems

The quality characteristics of an AI system can be evaluated based on a combination of quality characteristics from ISO 25010 and other quality characteristics. Some of the important characteristics from the AI testing perspective are:

- **Functional suitability**
    - Is one of the most important parameters for accepting a solution.
    - Functional correctness, completeness and appropriateness are expected with some sort of error estimation to quantitatively measure the system.
- **Reliability**
    - Availability of the system during normal operations.
    - Fault tolerance of the system should be high enough to handle corrupt data, incomplete, or irrelevant data without breaking down.
- **Performance efficiency**
    - Time behavior - how quickly the system responds to the demands made from it.
    - Resource utilization - Which and how many resources are used by the system to perform a function.
- **Maintainability**
    - Analyzability - The degree of effectiveness and efficiency with which it is possible to assess the impact on a product or system of an intended change to one or more of its parts, or to diagnose a product for deficiencies or causes of failures, or to identify parts to be modified. In the case of AI, analyzability also refers to the ability to be able to understand why the system took the decision that it took. Ideally, explainablity (or examinability) should be a separate quality characteristic for ISO 25010 for AI systems.
    - Testability - The degree with which the AI component supports testing in a given context. The higher the degree/the testability, the easier it is to find bugs by means of testing.

Additional important parameters are:

- **Complexity** – Time and space complexity.

- **Scalability** – The ability of the system to handle more load by adding additional resources to it.

- **Continuous learning** – The ability of the system to continuously learn from new data, especially from real time environment data.

2.3.2 Extended Quality Characteristics Specific to AI

The use of AI has required an extension of the standard quality characteristics.

An intelligent machine should also bear the following quality characteristics, apart from those mentioned in ISO25010. [TDA]

- Intelligent behavior – Intelligent behavior is the ability to comprehend or understand. It is basically a combination of reasoning, memory, imagination, and judgment; each of these faculties relies upon the others. Intelligence is a combination of cognitive skills and knowledge made evident by behaviors that are adaptive. The sub-characteristics are: Ability to learn, Improvisation, Transparency of choices, Collaboration and Natural interaction.

- Morality – Morality, in relation to AI, is about the principles concerning the distinction between right and wrong or good and bad behavior. The sub-characteristics are: Ethics, Privacy and Human friendliness.

- Personality – Personality is the combination of characteristics or qualities that form an individual's distinctive character. The sub- characteristics are: Mood, Empathy, Humor and Charisma.