NIST AI Technology Evaluation Test Specification
Quantum Dot Control

Version 1.0— May 29, 2026

Test Version:	1.0
Task:	Three-state Probability Estimation
Input:	Single Image and Text
Output:	Text
Task Champion(s):	Justyna P. Zwolak, Merritt Losert, Brian Weber, Zachery Merino
Dataset:	QDC Patches v1.0
Metric:	A cost function that corresponds to the mean squared error will be computed.

Introduction

Quantum dots are nanoscale semiconductor structures that can trap and control individual electrons as *charge islands*, with myriad applications, e.g., single-electron transistors, quantum computing, and medical imaging. Tuning quantum dot devices has long been considered an experimental "art," requiring expert intuition and time-intensive measurements. These types of trial-and-error and heuristic approaches do not scale and thus cannot be used to scale quantum technologies, and automated methods for quantum dot tuning have the potential to overcome these limitations. The task described in this test represents a basic ability for quantum dot control and, as models demonstrate their abilities and improve over time, additional, increasingly complex tasks in support of quantum dot control will be added to the test.

Task Definition

Charge stability diagrams are two-dimensional measurements that show how the electrical response of a quantum dot changes as voltages applied to control electrodes are varied. Given an image patch of a charged-stability diagram, the model must output a three-state probability vector that the image patch displays a double-dot, a single-dot or no dot.

Dataset Description

The test data consists of images representing small regions of quantum dot charge stability diagrams.

1. *Modalities:* Images
2. *Data creation/collection date(s):* April 2026
3. *Derived Size:* $\sim$ 0.5 GB (641 images)

4. *Reference labels:* A custom algorithm transforms the experimental images into wireframe charge stability diagrams, then a geometric-based technique utilizing well-established physics principles is used to determine posterior probability of each class given the diagram.

5. *Inter-annotator Agreement/Annotation and Data Quality Assessment:* A visual inspection of all labeled charge stability diagrams was performed independently by four domain experts, confirming agreement between the assigned labels.

6. *Other pertinent features:* None

Testing Protocol

For each image in the test dataset, ask the model to provide a 3-dimensional vector, where each dimension corresponds to the probability the quantum dot is in each of the three possible states: no dot, single dot, or double dot.

1. *Input:* One image and a fixed prompt
2. *Output:* 3-d vector
3. *Scoring procedure(s):* Get output from model, compare against ground truth, compute error rates and metric value

Metric and Score Reporting

The primary metric will be the mean of the mean squared error computed across dimensions.

$$C_{Primary} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{3}\frac{(output_i^j - label_i^j)^2}{3} \tag{1}$$

where $output_i^j$ is the j^{th} dimension of vector output by the model for image i and $label_i^j$ is the the j^{th} dimension of ground truth label vector for image i.

Uncertainty will be computed using bootstraping methods[1].

Change Logs

- Version 1; May 29, 2026; Initial Release

[1] Efron, Bradley. "Bootstrap methods: another look at the jackknife." Breakthroughs in statistics: Methodology and distribution. New York, NY: Springer New York, 1992. 569-593.