

AI and Robotics (AIR) Solution Validation Platform

Crystal LM Fok* and Eric KS Cheng

Hong Kong Science and Technology Parks Corporation, 5/F, Building 5E, Science Park East Avenue, Hong Kong Science Park, Hong Kong

Abstract

For Artificial Intelligence (AI) and Robotics (R) solutions to achieve mass adoption across industries, impartial performance evaluation to establish trustworthiness is fundamental, before the AIR technology field can progress to the governance stage with a set of industry-accepted best practices and enforceable regulatory frameworks. Currently, there is a disconnect between the solution seekers and solution providers of AIR technology due to different foci – solution providers are primarily concerned with the science and capabilities of the technology, while solution seekers are looking at outcomes. The technology is stuck in the bind of requiring regulation but not yet reaching the regulation stage due to the missing link of independent performance evaluation.

Hong Kong Science and Technology Parks Corporation has become the first in the world to establish an AIR Validation Platform as an independent third-party performance evaluation tool to address this missing link and help build the trustworthiness of AIR solutions. The platform uses a combination of Physical and Virtual labs based on a comprehensive set of consideration factors of relevance to solution seekers and solution providers to establish an objective set of performance metrics and build consensus for them.

A Physical Lab will be equipped with a range of hardware to enable AI and Robotics functional performance and safety testing. It will support solution providers to test and measure the actual performance parameters of a solution such as accuracy, repeatability and mobility. Then those parameters will be used to create a corresponding virtual AIR solution in a Virtual Lab together with different application scenarios provided by solution seekers, such as a hospital, hotel or shopping mall, as a digital twin. Limitless combinations of variables can be added to the Virtual Lab to ascertain a solution's potential performance in varied settings and application scenarios of any industry.

This paper on the AIR Validation Platform will articulate the challenge of establishing trust for AIR technology and explain in detail the first-in-the-market solution that covers all three critical parts of building trustworthiness – performance metrics, physical testing and virtual testing. The paper will also share initial results from two pilot studies and describe the platform roadmap for building trust, encouraging mass adoption and driving the market dynamics that will lead to global standards for regulation.

Keywords: Artificial intelligence; Robotics; AI Algorithms

*Corresponding author:

Crystal LM Fok, Hong Kong Science and Technology Parks Corporation, Science Park East Avenue, Hong Kong Science Park, Hong Kong.

✉ crystal.fok@hkstp.org

Citation: Crystal LM Fok, Eric KS Cheng (2020) AI and Robotics (AIR) Solution Validation Platform Vol.8 No.3: 55 10.36648/2349-3917.8.3.55

Received: August 07, 2020; **Accepted:** August 22, 2020; **Published:** September 02, 2020

Overview of the AIR Validation Platform

Bridging a gap in the AIR innovation journey

The roadmap for innovation in Artificial Intelligence and Robotics (AIR) begins with promoting thought leadership to instigate change and concludes when the technology field has arrived at a consensus for well-defined standards to inform regulations and establish governance [1]. A major gap in the journey is the absence of a reliable and objective means to evaluate the performance and trustworthiness of an AIR solution by an impartial third-party.

Existing simulation or validation models are either driven by product development testing from the perspective of solution providers only, or if they address the needs of solutions seekers, they do not take into account the actual performance of the solution in the physical environment that may inevitably have multiple variables emerging spontaneously [2].

The absence of a trusted neutral platform to validate the performance of an AIR solution in varied actual application scenarios, and answering the exact needs of different groups of potential users, is hindering technology adoption, thus slowing the long-term development of AIR technology without mass adoption support.

The AIR Validation Platform pioneered by Hong Kong Science and Technology Parks Corporation (HKSTP) intends to bridge this gap in the AIR innovation roadmap. The platform serves as an objective and trustworthy third-party validation platform, primarily driven by solution seekers' needs, to examine and compare the performance of AIR solutions in the same category. By addressing solution seekers' requirements and gaining users' confidence, the platform seeks to encourage wider adoption of AIR solutions.

Input and output

The AIR Validation Platform is the first in the world to apply performance matrices, physical testing and virtual testing in one cohesive validation mechanism to identify the most suitable AIR solution for performing required functions [3].

A universal Performance Matrix applicable to a specific category of AIR solutions defines the framework for validation. The criteria in the matrices are inclusive of all consideration factors of relevance to solution seekers and solution providers. It is a critical step in the validation process for establishing a user-developer consensus on the performance evaluation standards from the outset.

Robots to be tested are operated in a Physical Lab to collect their actual performance parameters, which are inputted into a Virtual Lab where the corresponding digital twins of the robots are operated with full visualization. The Virtual Lab can be visualized in an infinite number of iterations to test the robots' performance in different application scenarios. All assumptions about how a robot will perform in a varied virtual setting are verifiable by taking the robots back to the Physical Lab to confirm actual

performance parameters. Integrating hardware in the loop is the missing component in the current visualization validation model in the market [4].

The validation of AI solutions is a similar process, with actual sample datasets being provided by solution seekers (instead of the hardware in the loop) for verification in the performance testing based on a set of AI algorithms.

On completion of the virtual to physical to virtual testing process, all performance data is inputted into the matrix for calculation. The score for each criterion in the Performance Matrix is calculated based on different weightings determined by the priorities and preferences of individual solution seekers. A final report evaluating different solutions based on the same set of criteria will give solution seekers a clear picture about how to select the right solution for their intended purposes.

Platform architecture

The AIR Validation Platform is structured for a three-stage workflow (Figure 1).

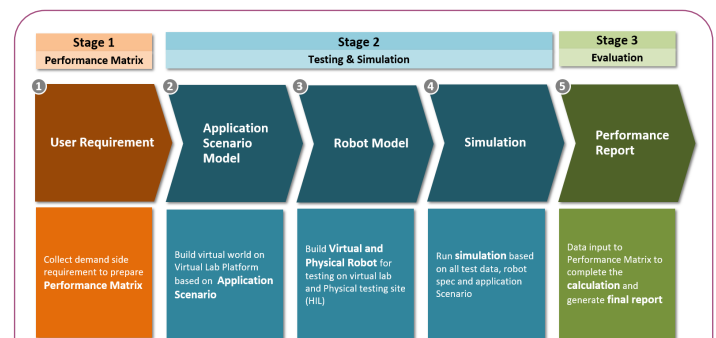


Figure 1: AIR validation platform general workflow.

The first stage focuses on gauging user requirements to develop a Performance Matrix that is universally applicable to similar solutions and accepted by both solution seekers and solution providers [5].

Stage two is testing and simulation. Physical robots are tested in a Physical Lab to collect performance parameter data. To overcome physical limitations in modifying a space to test performance in varied settings, a 3D virtual setting mirroring the actual physical site for operating the robots is built in the Virtual Lab.

The virtual setting can be arranged in infinite iterations, by means of adding and subtracting objects, assets, traffic flow and every other item that may possibly be present in that 3D structure. Digital twins of the devices to be tested in the Virtual Lab are dropped into the setting to simulate performances in different conditions.

Thus far, it is not that much different to creating a virtual game setting. The performance of a device in a virtual setting is based on a set of assumptions. To verify that the assumptions are correct, the hardware can be brought into the testing loop and physically tested again to ensure that the virtual simulations accurately represent real-life performance parameters.

In the final stage, unambiguous evaluation results will be

produced through a video visualization of the physical and virtual tests and an operational analysis, the scoring of the robots for each criterion in the Performance Matrix, and a full dataset showing every aspect of the test performances in the Physical Lab and the Virtual Lab.

The test and simulation process for AIR validation can be applied to the deployment of any hardware solutions, such as the optimal mapping of Wi-Fi stations to achieve ideal coverage.

The testing of AI solutions is similar, in that the initial step is defining the Performance Matrix based on solution seekers' requirements. But instead of using hardware for testing, solution seekers will upload their sample datasets to the platform's API server, where solution providers will also upload their AI algorithms to address the solution seekers' requirements. Performance testing will be scheduled through resource controllers and then evaluated by a validation engine working on a set of pre-defined algorithms based on the Performance Matrix. The test data is stored in the API server for outputting to the final report that matches the optimal solution to the potential user's needs [6].

Case Studies

Robotics: Disinfection robots

The Covid-19 pandemic in early 2020 created enormous demand for disinfection robots. A dozen models emerged in the market using different disinfection modes – UV, spraying or fogging. Any claims by robotics providers about the performance of the robots in different settings – be it an office building, hospital, hotel, shopping mall or other large facility – or in relation to their disinfection efficiency, ability to steer clear of objects or people or any other performance parameters could not be substantiated, so users were unable to make informed choices.

Through the AIR Validation Platform, we established a Performance Matrix based on the attributes solution seekers desire, relating to business considerations, performance and quality, operation, safety and technical specifications, and determined from which channel to collect the performance feedback – the solution adopter, the robot operator, the robot developer, third-party testing by other professional labs or from the platform's Virtual Lab or Physical Lab.

The robots were first tested in the Physical Lab to gather performance data to be fed into the virtual simulations. Digital twins of the tested robots were created to operate in the Virtual Lab simulating actual application scenarios with multiple variables, such as operating around different obstacles or human traffic. The performance data of the robots' simulated operations is tested again in the Physical Lab for verification. A multi-camera motion capture system was set up to record the performance, and algorithms were developed to calculate the variance between the performance parameter assumptions in the simulated test and the actual performance capabilities [7].

To evaluate the performance criteria related to safety and technical specifications, third-party professional labs were engaged. For example, a commercial test lab was engaged to

measure sanitization effectiveness, by collecting environmental sample data before and after the disinfection. The AIR Validation Platform also consulted an expert group with extensive domain knowledge to provide professional advice on the most effective frameworks for validation.

The outcome was presented with the visualization of the robots' functioning within a building and operational analyses simulating environments with heavy human traffic to compare the efficiency difference in deploying multiple robots. Based on the solution seeker's needs, we could make confident recommendations on which type of robots, and the optimal number of robots, to be deployed within specified time frames.

The results also provided a valuable reference for the robot providers about how to improve their products on an ongoing basis.

The virtual visualization of the robots' performance gave the solution seeker confidence for adoption, and the knowledge to prepare tenders for disinfection robots. The process enhanced user confidence, which is conducive to wider scale adoption.

The universal set of performance parameters can be adopted for testing all disinfection robots that enter the market in the future. Potential adopters of disinfection robots can reference the Performance Matrix, and score robots by giving a higher weighting to the attributes they consider most important.

The Virtual Lab built for testing disinfection robots can be adapted for testing other AIR solutions. The potential usage of the Virtual Lab is infinite given the flexibility to drag and drop new elements from the asset library to tailor-create a new scenario. By allowing unlimited iterations, The Virtual Lab reduces the need for space in the Physical Lab. In this case, a 1,000 square foot space was adequate for physical testing.

Through the validation exercise involving multiple disinfection robots, we can build a substantial database of solutions to help solution seekers make informed decisions and smart choices, and help the industry achieve best practice by setting performance standards defined by the Performance Matrices.

AI: Facial recognition access control

AI-powered facial recognition is multi-functional – for unlocking phones, identity authentication, and access control, finding missing persons or other uses. The requirements for facial recognition outcomes could be vastly different for users in different industries. For example, shopping malls may use facial recognition for early detection and recognition of loyal customers and shoplifters, while airports may use facial recognition to automate access control for a large number of passengers and visitors, with minimal risk of letting trespassers through [8].

The AIR Validation Platform can provide neutral and fair testing to help solution seekers turn requirements into measurable metrics and prepare high-quality data, which helps them choose the best solution for their requirements with a standardized set of judging criteria and identical computing resources.

The Performance Matrix for facial recognition solutions, setting industry-accepted standards for evaluation, takes into account the following attributes, calculated by a range of formulas and modeling:

- **Classification accuracy** – measures a solution's efficiency for correct identification, based on the ratio between the number of correct predictions and the total number of predictions made.
- **Logarithmic loss** – applies penalty points in the to deduct scores for false identifications.
- **Confusion matrix** – calculates true positives, true negatives, false positives and false negatives.
- **Area under curve** – Identifies the true positive rate and false positive rate.
- **F1 score** – rates the robustness of the classification of instances.
- **Mean absolute error** – determines the average of the differences between the original values and the predicted values.
- **Mean squared error** – Enables the model to focus on larger errors, the effect of which becomes more pronounced.

The AI validation process primarily involves defining the Performance Matrix and helping solution seekers collect high-quality and robust sample datasets that have minimal biases and do not weigh heavily on extremes. A solution's performance is measured on each criterion in the matrix. Extra weighting is assigned to the criteria that the solution seeker considers of primary importance. For example, in the case of an airport, logarithmic loss assumes a higher weighting because the risk of admitting a trespasser is considered most costly. The final analysis offers solution seekers a clear recommendation for the solution that best meets their needs.

Conclusion

The AIR Validation Platform is capable of delivering measurable benefits to all parties involved in the AIR technology development process.

Impartial validation testing connecting physical and virtual testing with hardware and authentic operational sample data in the loop yields objective facts as a trustworthy reference for solutions seekers to compare solutions in the market and adopt a best-fit solution with confidence. Solution providers can reference performance data of rival solutions in the market to strive for improvement. The Virtual Labs can be easily modified for testing an unlimited variety of solutions.

Confidence in AIR solutions encourages mass adoption, which in turn stimulates further solution development, and creates a virtuous circle for accelerating best practice and building the groundwork for setting standards as the framework guiding regulations and governance.

HKSTP is in the best position to pioneer the AIR Validation

Platform. Already an enabler with a diverse range of incubation and acceleration programs, HKSTP also serves an active role in connecting solution providers and solution adopters through a multitude of matching and industry-connect programs. As a statutory body tasked to promote I&T development in Hong Kong, its impartiality is unquestionable.

There are several pilot projects for building the Virtual Lab in the pipeline, including disinfection robots, inbound logistics robots, construction welding/painting robots and curtain wall cleaning robots.

When a technology has been commonly used, the standard timeframe for progressing to governance is five to seven years. The implementation of the AIR Validation Platform will facilitate and expedite the process towards establishing governance within the next decade.

References

1. Bostelman R, Falco J, Hong T, Messina E (2016) Performance Measurements of Motion Capture Systems used for AGV and Robot Arm Evaluation.
2. Guinn B (2019) Validating A Decontaminations Protocol Utilization Ionized Hydrogen Peroxide.
3. Notes on Evaluating Face Recognition Software (2020).
4. Mishra A, (2018) Metrics to Evaluate your Machine Learning Algorithm. From Towards Data Science.
5. Chen K, Lu D, Chen Y (2014) The intelligent techniques in robot Keija—The champion of RoboCup@Home 2014. RoboCup 2014: Robot World Cup XVIII 8992: 130–141.
6. Chen X, Xie J, Ji J (2012) Toward open knowledge enabling for human–robot interaction. Human–Robot Int 1: 100–117.
7. Krishnamurthy J, Kollar T (2013) Jointly learning to parse and perceive: Connecting natural language to the physical world. Transactions of the Association for Computational Linguistics 1: 193–206.
8. Aggarwal JK, Ryoo MS (2011) Human activity analysis: A review. ACM Computing Surveys 43: 16.