



# Pioneering a path to robust **algorithm testing**



**Mitchell Sipus**  
Director of Product, CalypsoAI

Machine learning is only valuable for the government if it can be tested, validated and trusted

**A**RTIFICIAL INTELLIGENCE HOLDS great promise for improving government services and operations. However, unlike traditional technologies, there are no universal benchmarks for how machine learning algorithms should function, and every algorithm has trade-offs in speed, efficiency, accuracy or resilience to adversarial attacks.

Let's say a program office wants to buy an algorithm to understand and better inform investments in public housing. Such an algorithm can make sure that every dollar spent is going to the best use. But what if the data from one part of the country is a little different? Or what if an unusual situation happens, such as a drought slowly displacing thousands of people? These subtle changes might result in untrustworthy results. Performance scores will remain high, and yet millions of dollars will be misspent and people will be worse off.

Every day, new weaknesses are discovered as algorithms are revealed to generate racist, sexist or other undesirable behaviors. How do agencies know if an algorithm is reliable? How can they compare algorithms from different vendors? These problems are slowing federal AI programs to a snail's pace.

## Assessing resiliency and boosting confidence

To accelerate federal missions, CalypsoAI is pioneering new ways of testing algorithms so that users can rapidly understand, assess and trust them. CalypsoAI's product, VESPR Validate, will run a battery of tests and deployment simulations. Will an algorithm work in the snow? What happens if it is hacked? These tests are rooted in current research efforts at top universities but are hardened for commercial demands and delivered as a federally compliant solution.

For the first time, mission owners can simply upload an algorithm and a batch of test data to see an easy-to-read report on when the algorithm succeeds and when it fails.

## Powering mission outcomes

As a result of the tests we've developed, leaders at the Air Force, the Department of Homeland Security and the Coast Guard are changing their way of thinking about AI acquisitions, their role as

leaders in developing AI programs and how to measure the impact of those efforts. For the first time, they have a roadmap to quickly procure the capabilities they need and assess models across vendors while understanding the risks and thresholds of their models to better protect the interests of the American people.

Our goal is to give government agencies access to reliable machine learning models and enhance their trust in the algorithms that can power mission outcomes. ■

Mitchell Sipus is director of product at CalypsoAI.

CALYPSOAI

87% of Data Science Projects Never Make it into Production

- Lack of Trust
- Lack of Understanding
- Lack of Leadership Buy-In
- Lack of Robustness & Security

With CalypsoAI, leaders can be confident their AI solutions are ready to be fielded safely & securely across business & society.

[Learn More](#)

[www.calypsoai.com](https://www.calypsoai.com)