October 6, 2021

CalypsoAI's Software to Facilitate the Validation, Monitoring, and Security of Enterprise AI Platforms

AI Test, Evaluation, Verification and Validation

Andrew Spage
andrew@calypsoai.com
(703) 629-2978

CALYPSOAI

So What? Who Cares?

S&T AI/ML Strategic Plan

GOAL 1: Drive Next-Generation AI/ML Technologies for Cross-Cutting Homeland Security Capabilities

Mission: Research and develop AI/ML testing tools

- Accelerate *Trusted AI* development and adoption, realize the benefits of automation

Problems: Insufficient AI/ML performance verification and validation

Solution: 3rd party AI Test & Evaluation

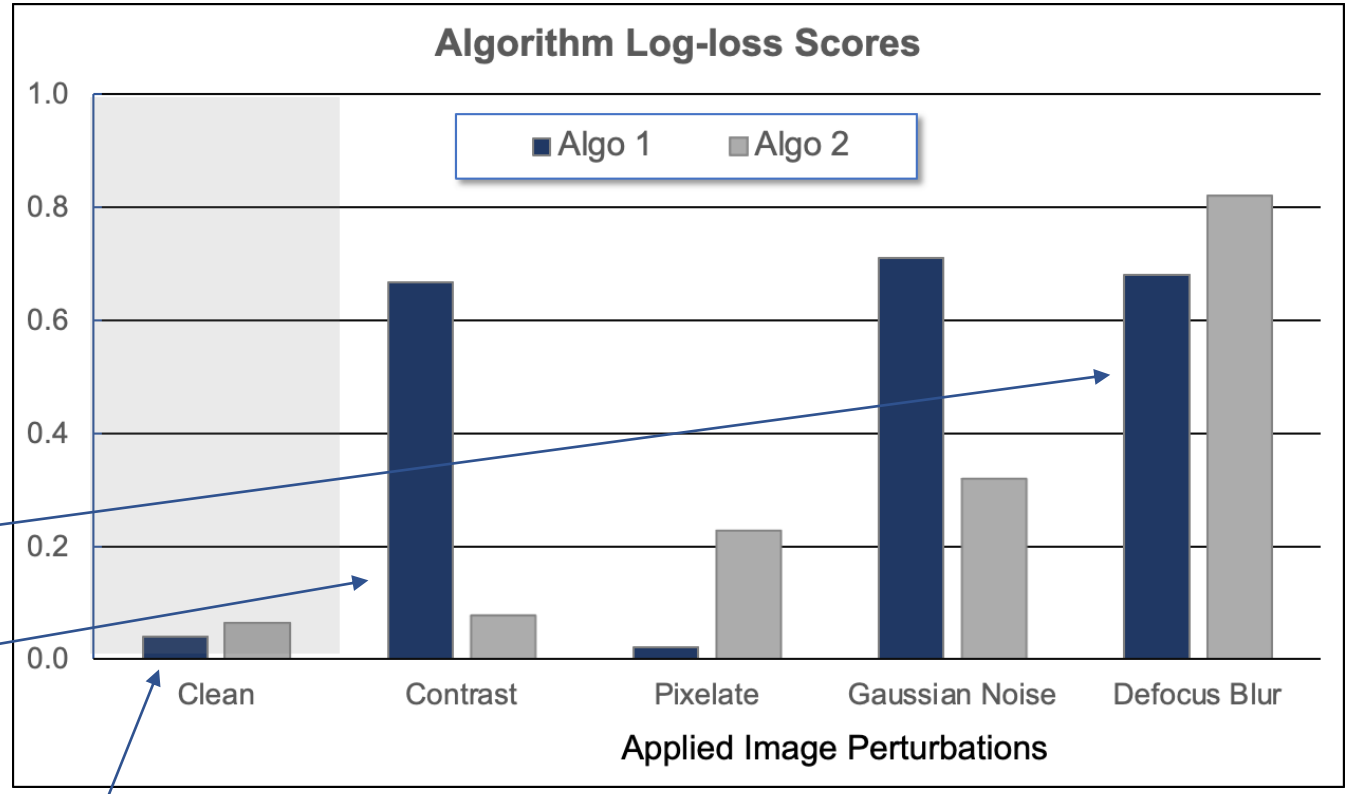
- Quantify performance variability to understand it, control it, and minimize it

Results: AI/ML model performance measures aligned to real-world screening scenarios and operational goals

Use Case – Testing Beyond Accuracy

Testing Automated Threat Detection (ATD) algorithms

- Baseline algorithms performed with high levels of accuracy against test data
- Robustness testing showed degraded run-time scans caused algorithms to fail precipitously



Higher Log Loss indicates reduced performance

Algorithms fail at varying rates depending on the perturbation applied

Base Case

Use Case - Conclusions

- Testing revealed that ATD algorithms, while accurate, did not tolerate degraded imagery
  - Baseline algorithms performed with high levels of accuracy against pristine test data
  - Robustness testing showed degraded run-time scans caused algorithms to fail sharply
- Testing indicated an issue, and potential paths to mitigate it
  - Robustness could be mitigated in training data development
  - Implementing image quality tests in run-time deployment mitigate risk of model failure

Considerations for the Way Ahead

Curate training data to match operational conditions and risks

- Include run-time inputs and sensor performance characteristics

Exploit the physical world

- Apply physical science expertise in training data development (subject matter expertise)
- Partnerships are essential (physical scientists, data scientists, hardware manufacturers, software developers, platform engineers, operators, standards)

Manage variability in the operational environment

- Deployed AIs operate as a component in a system, integration will impact performance

Define “trust” in mission outcomes

- Test for conditions that threaten bad outcomes and contain them

Questions?

Andrew Spage
Director of Solutions Architecture
CalypsoAI
(703) 629-2978
andrew@calypsoai.com