CALYPSOAI

Empowering your AI Missions through Independent Validation & Testing

An Introduction to CalypsoAI

CalypsoAI was established to solve the big challenges facing AI deployment today

AI Experience coming from the US National Security community, and key investors who invest across AI categories

Customers and Strategic Relationships:

U.S. AIR FORCE

NATIONAL AIR & SPACE INTELLIGENCE CENTER

U.S. DEPARTMENT OF HOMELAND SECURITY S&T

DEVCOM ARMAMENTS CENTER

TEST RESOURCE MANAGEMENT CENTER DEPARTMENT OF DEFENSE

CDAO

SAIC

pwc

ECS

Founded: 2018

Locations: San Mateo, CA and Virginia

Stage: Series A; Series B exp. late 2022

Investors: Paladin Capital, Lightspeed Ventures, 8VC, and Lockheed Martin Ventures.

Contract Vehicles: JAIC T&E BPA, NASA SEWP, GSA, Army CHESS ITES-SW2, other contracts available upon request

Deployment Options: Built on Kubernetes, we support containerized deployment and are hardware/platform agnostic. Available via AWS GovCloud and on-prem

Named Gartner®
Cool Vendor
AI Core
Technologies 2022

Gartner
Cool Vendors in
Artificial Intelligence
2022
CALYPSOAI

Billions Spent – Why Aren't we Deploying More AI?

AI is the most critical technology of the 21st century but organizations are struggling to operationalize their AI models

Erroneous outcomes, lack of standardized model testing, and lack of trust in AI/ML means models are not deployed into production

These issues are exacerbated when:

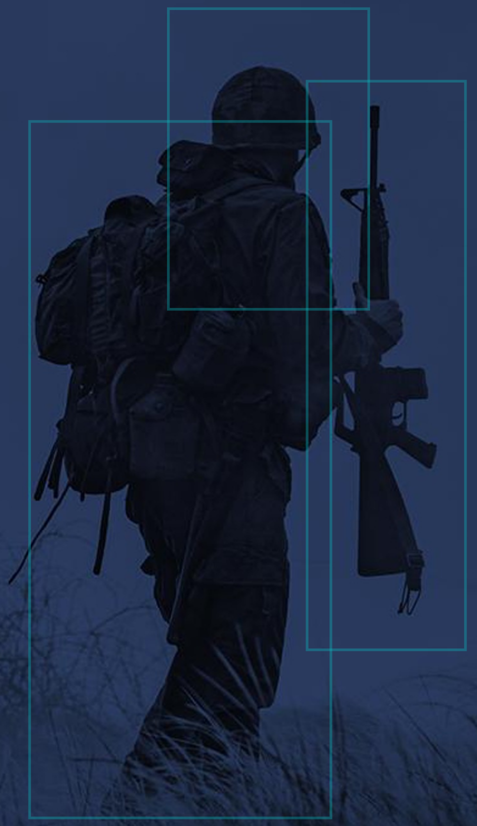
- Models are trained on limited data that does not represent the deployment environment
- Models are not developed to withstand rapid change in environments or real-world conditions
- Models are at risk of strategic adversarial noise injection
- Model test and evaluation is not automated
- There is a lack of model version control
- There is a lack of model performance standardization

CALYPSOAI

VESPR Validate

The solution that builds Trust in your AI adoption by independently validating and testing your AI/ML models.

Confidently enable and accelerate your Mission!



Our Product: VESPR Validate

Ensuring your AI/ML can achieve organizational goals, securely, in real-world conditions.

Stress Test Real-World Performance

Utilizing 3D maps, gaming engines, physics-based simulations, and quantified noise distributions mirroring real-world data gaps we test models in adverse environments to provide confidence of accurate performance in operational environments. These include:

Weather Conditions | Blur | Brightness | Defocus

Inversion / Privacy Testing

Performing rapid systematic attacks on the model to inference sensitive training data, we determine if this data is secure.

Adversarial Security

We use cutting edge adversarial attacks on the model to trigger model failure utilizing the Minimal Attack Surface to test model vulnerability to adversarial image attacks.

Use via GUI or Integrate via API/SDK Toolkit

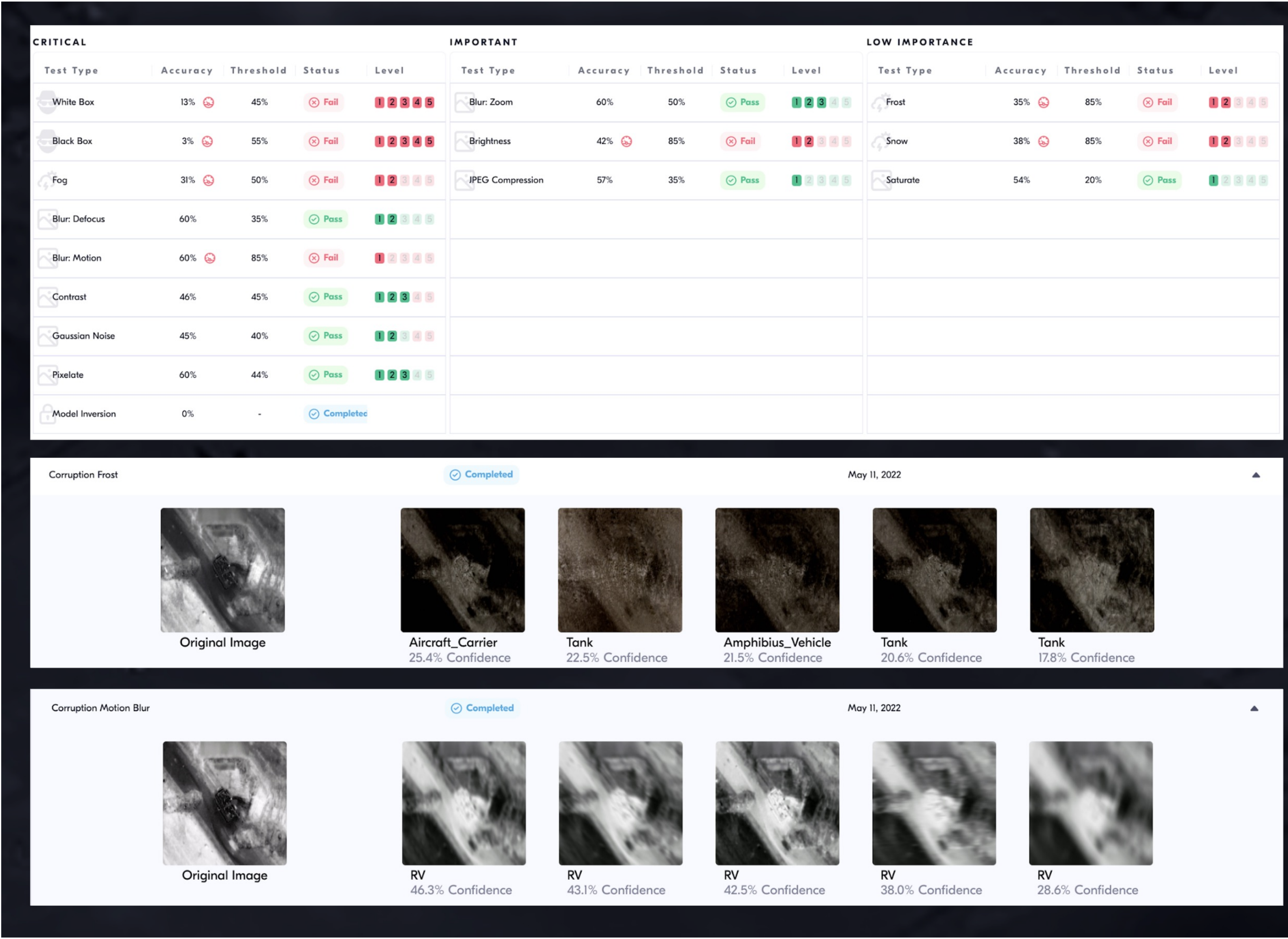
CALYPSOAI

Case Study

Automated Target Recognition (ATR) from MQ-9 Full Motion Video

Data: Full Motion Video (FMV)
Source: MQ-9 Reaper
Data Type: Infrared
Target: Tank
Model Type: Pytorch Classifier
Number of Classes: 11

Vendor Model Performance Metrics
F1 Score .58
Global Accuracy .59
Precision .61
Recall .59



CRITICAL

Test Type	Accuracy	Threshold	Status	Level
White Box	13%	45%	Fail	1 2 3 4 5
Black Box	3%	55%	Fail	1 2 3 4 5
Fog	31%	50%	Fail	1 2 3 4 5
Blur: Defocus	60%	35%	Pass	1 2 3 4 5
Blur: Motion	60%	85%	Fail	1 2 3 4 5
Contrast	46%	45%	Pass	1 2 3 4 5
Gaussian Noise	45%	40%	Pass	1 2 3 4 5
Pixelate	60%	44%	Pass	1 2 3 4 5
Model Inversion	0%	-	Completec	

IMPORTANT

Test Type	Accuracy	Threshold	Status	Level
Blur: Zoom	60%	50%	Pass	1 2 3 4 5
Brightness	42%	85%	Fail	1 2 3 4 5
JPEG Compression	57%	35%	Pass	1 2 3 4 5

LOW IMPORTANCE

Test Type	Accuracy	Threshold	Status	Level
Frost	35%	85%	Fail	1 2 3 4 5
Snow	38%	85%	Fail	1 2 3 4 5
Saturate	54%	20%	Pass	1 2 3 4 5

Corruption Frost ✓ Completed May 11, 2022

Original Image	Aircraft_Carrier 25.4% Confidence	Tank 22.5% Confidence	Amphibius_Vehicle 21.5% Confidence	Tank 20.6% Confidence	Tank 17.8% Confidence

Corruption Motion Blur ✓ Completed May 11, 2022

Original Image	RV 46.3% Confidence	RV 43.1% Confidence	RV 42.5% Confidence	RV 38.0% Confidence	RV 28.6% Confidence

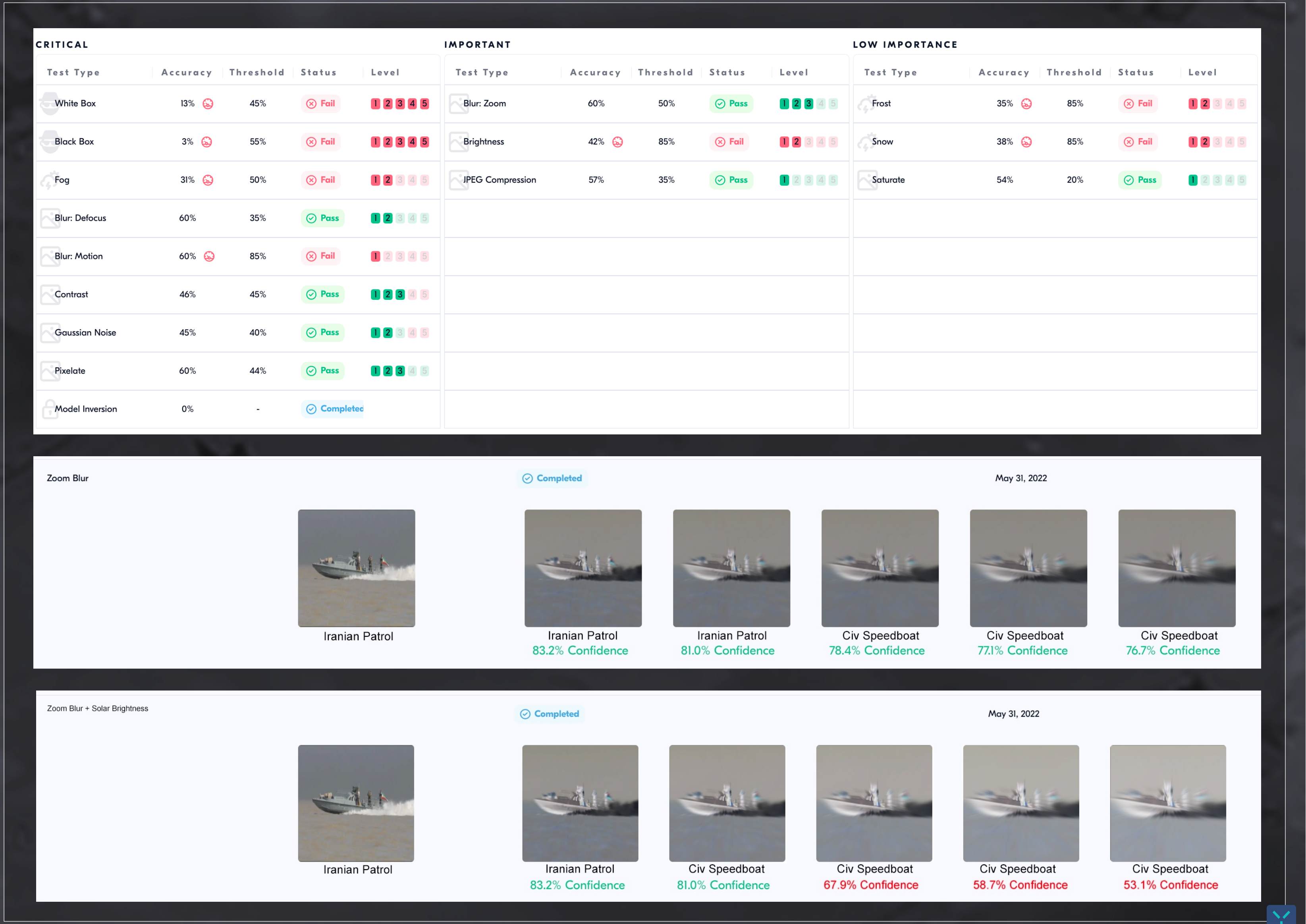
Case Study

Sail Drone, Automated Identification of Iranian State Actors, Faris Island

Location: 27.9900° N, 50.1700° E
Near: Farsi Island, W, NW

Data: Full Motion Video (FMV)
Source: Unmanned Sail Drone
Data Type: Optical, Infrared
Target: Iranian Patrols, Small Boats
Model Type: Pytorch ResNet V1.5
Number of Object Classes: 11

Vendor: Model Performance Metrics
F1 Score .87
Global Accuracy .86
Precision .9
Recall .89



CRITICAL

Test Type	Accuracy	Threshold	Status	Level
White Box	13%	45%	⊗ Fail	**1** **2** **3** **4** **5**
Black Box	3%	55%	⊗ Fail	**1** **2** **3** **4** **5**
Fog	31%	50%	⊗ Fail	**1** **2** 3 4 5
Blur: Defocus	60%	35%	✓ Pass	**1** **2** 3 4 5
Blur: Motion	60%	85%	⊗ Fail	**1** 2 3 4 5
Contrast	46%	45%	✓ Pass	**1** **2** **3** 4 5
Gaussian Noise	45%	40%	✓ Pass	**1** **2** 3 4 5
Pixelate	60%	44%	✓ Pass	**1** **2** **3** 4 5
Model Inversion	0%	-	✓ Completed	

IMPORTANT

Test Type	Accuracy	Threshold	Status	Level
Blur: Zoom	60%	50%	✓ Pass	**1** **2** **3** 4 5
Brightness	42%	85%	⊗ Fail	**1** **2** 3 4 5
JPEG Compression	57%	35%	✓ Pass	**1** 2 3 4 5

LOW IMPORTANCE

Test Type	Accuracy	Threshold	Status	Level
Frost	35%	85%	⊗ Fail	**1** **2** 3 4 5
Snow	38%	85%	⊗ Fail	**1** **2** 3 4 5
Saturate	54%	20%	✓ Pass	**1** 2 3 4 5

Zoom Blur ✓ Completed May 31, 2022

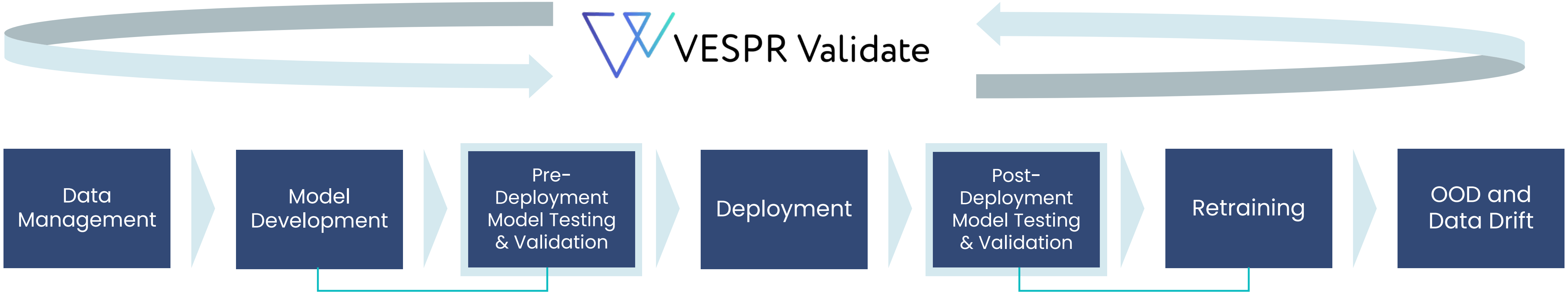
Iranian Patrol	Iranian Patrol 83.2% Confidence	Iranian Patrol 81.0% Confidence	Civ Speedboat 78.4% Confidence	Civ Speedboat 77.1% Confidence	Civ Speedboat 76.7% Confidence

Zoom Blur + Solar Brightness ✓ Completed May 31, 2022

Iranian Patrol	Iranian Patrol 83.2% Confidence	Civ Speedboat 81.0% Confidence	Civ Speedboat 67.9% Confidence	Civ Speedboat 58.7% Confidence	Civ Speedboat 53.1% Confidence

MLOps Pipeline

Integrating AI/ML Testing and Validation Through CI/CD as a Core Practice



VESPR Validate works across the MLOps pipeline and can easily integrate with MLOps tools such as but not limited to Azure Machine Learning, Scalabel, DataRobot, Dataiku, Arize AI, and many more.

Data Management Solution + Model Development / Monitoring + CalypsoAI's VESPR Validate = End-to-End MLOps Solution

VESPR Validate offers critical automated Test, Evaluation, Validation & Verification (TEVV) components to enable organizations to create a robust MLOps platform that ensures models function correctly in operational environments characterized by rapid change, adversarial activity, and varying mission profiles

Thank You

CALYPSOAI