# Use of Simulation to train AI for swarm based underwater behavior – Lessons Learned from Talisman Sabre 2025

| Dr. Peter Drewes | Dr. Richard Choroszucha | Anthony Cruz |
|---|---|---|
| SAIC | VATN Systems | OASD . OUSD R&E |
| Reston, Va | Bristol, RI | Washington DC |
| Peter.Drewes@saic.com | rc@vatnsystems.com | Anthony.s.cruz8.ctr@mail.mil |

## ABSTRACT

Recent trends in unmanned systems have indicated the need for large swarms of unmanned systems to overcome adversary capabilities. Part of the problem is the need for collaborative sensors and behaviors working together. In support of this exploration, the Office of Secretary of Defense is exploring Resilient and Autonomous Artificial Intelligence Technology (RAAIT) trials to collaborate data feeds in the experimentation process. The current exploration is through a US / INDOPACOM joint experiment called Talisman Sabre 2025. This Live / Virtual / Constructive training environment is used to test science and technology capabilities integrated with military exercises.

This paper covers the lesson's learned using collaborative autonomous underwater vehicles (AUVs) that were trained using the Advanced Framework for Simulation (AFSIM) connected to an AI engine to create tactical and collaborative behaviors that were operated in simulated and live environments. These behaviors guided the AUVs as a team to behave in a counter underwater vehicle solution integrated into the RAAIT solution for the purposes of detecting and tracking adversary underwater assets to provide early warning for defense of high value underwater assets. This research demonstrates the value of using Modeling and Simulation and AI/ML for reducing time to create behaviors and deploy them on live unmanned assets through the experimentation process. This experimentation process forms the efficiency of the simulation result versus the real-world capabilities.

## ABOUT THE AUTHORS

**Dr. Peter Drewes** – is the Chief Technology Officer for the business development group at SAIC. He received his Ph.D. from the University of Central Florida. He has concentrated his research on performance monitoring in real-time systems. Currently, Dr. Drewes is a technical resource for SAIC, focusing on the areas of robotic systems and connections to Live, Virtual and Constructive simulation environments.

**Dr Richard Choroszucha -** Dr. Richard Choroszucha is a Principal Engineer at Vatn Systems. He holds a Ph.D. in Naval Architecture and Marine Engineering from the University of Michigan - Ann Arbor. His work has included simultaneous localization and mapping, adaptive control, collaborative autonomy with humans-in-the-loop, coupling AI/ML with Guidance and Navigation systems.

**Anthony Cruz -** is an International Engagements Analyst supporting OUSD R&E Trusted AI & Autonomy. He received his MBA from Yonsei University. His work is focused on connecting systems together across multiple domains to explore fields of autonomy and artificial intelligence, military experimentation, financial investment, defense industry sales, and civil society development.

# Use of Simulation to train AI for swarm based underwater behavior – Lessons Learned from Talisman Sabre 2025


**Dr. Peter Drewes**
**SAIC**
**Reston, Va**
**Peter.Drewes@saic.com**

**Dr. Richard Choroszucha**
**VATN Systems**
**Bristol, RI**
**rc@vatnsystems.com**

**Anthony Cruz**
**OASD . OUSD R&E**
**Washington DC**
**Anthony.s.cruz8.ctr@mail.mil**


## INTRODUCTION

In future maritime operations, the proliferation of unmanned underwater vehicles (UUVs) by both state and non-state actors will require the Navy to field agile, scalable counter-UUV capabilities that can operate independently or as part of a larger force structure. A collaborative UUV group, trained to autonomously detect, classify, track, and neutralize adversarial UUVs, would operate as an integrated layer within the Navy's existing undersea warfare (USW) architecture. These autonomous teams would provide persistent underwater surveillance, threat neutralization, and cueing support to manned platforms such as submarines, destroyers, and maritime patrol aircraft, acting as both an early-warning sensor grid and a rapid response mechanism in denied environments.

This counter-UUV capability would be seamlessly integrated into distributed maritime operations (DMO) and the Navy's evolving unmanned campaign plan, ensuring flexible command and control (C2) at both the tactical and operational levels. UUV teams could operate under mission command principles, executing assigned objectives autonomously while remaining networked for periodic updates and coordination with broader task groups. In contested communication environments, autonomy would allow UUVs to continue executing mission objectives with minimal external input. Integration with international partners through frameworks such as AUKUS would enable coalition forces to deploy interoperable UUV packages, share underwater situational awareness data, and collaboratively defend critical sea lines of communication, chokepoints, and high-value units.

An appropriate experiment was selected to test this concept – Talisman Sabre 2025 (TS-25). TS-25 is a large-scale, biennial military exercise primarily involving the United States and Australia. It's designed to strengthen partnerships, enhance interoperability, and improve the combined military capabilities of participating nations in the Indo-Pacific region. The exercise includes a wide range of activities across various domains, including land, air, sea, cyber, and space. There are two portions of the exercise, those between the military organizations, and a Science and Technology (S&T) extension to mature capabilities during the exercise.

The SAIC team was invited to join the Office of the Secretary of Defense's TS-25 S&T activity under the Resilient and Autonomous Artificial Intelligence Technology (RAAIT) initiative. The goal of the RAAIT trials is to develop an artificial intelligence ecosystem that will one day enable the three partner nations to share data for operational success in contested environments. For the TS-25 portion, there will be unmanned air, ground, surface and subsurface assets providing information and taking taskings to support the Opposing Forces (OpFor) during the experimentation. The assets are used to provide additional intelligence and simulated effects to slow down Marine Corp elements (US and Coalition Forces) executing their operational mission.

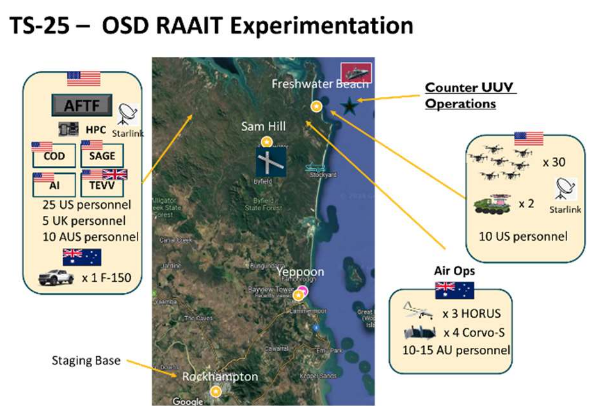


**Figure 1 – TS-25 RAAIT Operational Area**

Our portion of the experimentation support is a group of Unmanned Underwater Vehicles (UUV) performing a counter exercise (CUUV). The top-level concept is that the group of UUVs will be deployed from an Unmanned Surface Vehicle (USV) or support boat, then patrol the waterway in a coordinated manner. When an enemy underwater vehicle is detected, the group decides how best to intercept and track the vehicle while maintaining patrol coverage and reporting back to the Command and Control (C2) Environment. An USV will be used to provide mobile surface communications to the C2 centers and messages to/from the UUV fleet.

**Concept of Operations**

A fleet of four UUVs are deployed from the USV/support boat. Once launched, they form up on the surface and then descend to an appropriate depth as a group to begin the patrol mission. This can be considered a large box pattern in formation. There are several formation possibilities for the grouping. The reverse V formation like a flock of birds is the logical first choice as it provides additional possibilities of the vehicles seeing each other. As part of the experimentation process will be the theoretical formation efficiency versus the real-world results. Depending on the horizontal distance between the UUVs, the Forward-Looking Sonar (FLS) may interfere with each other as they each use the same frequency.

As the patrol UUV group descends to the operational depth (target of 15-30 meters), a target or enemy UUV is launched from another position and begins unknown motion into the patrol area. When one of the UUVs detects the enemy UUV, this is fed through the group behavior system to decide on which play should be run. For example, if the group is heading east, and an enemy vehicle is detected heading south by the left (or most northern vehicle), it might make sense for the most south vehicle to break off and chase even though it can't see it. Another ConOp is the one who finds it chases the target, while the others re-form around the gap left by the chasing vehicle. Depending on the speed and separation, it may require three-dimensional re-routing of the group to allow the chase UUV to move into position quickly without impacting its team mates.

Once the play has been called and the UUV moves to intercept and follow the enemy UUV, the others re-group and continue the patrol in looking for additional targets. A USV will be traveling above the UUVs and stay with the grouping to minimize communications distance for the overall grouping. The follower UUV will report its and the targets position as it moves in its new direction. As a follow on, the involved Naval fleet or C2 center could order the enemy UUV neutralized. However, the primary goal is to find the adversary, provide a fix of its location and leave what to do with the adversary to the C2 environment.

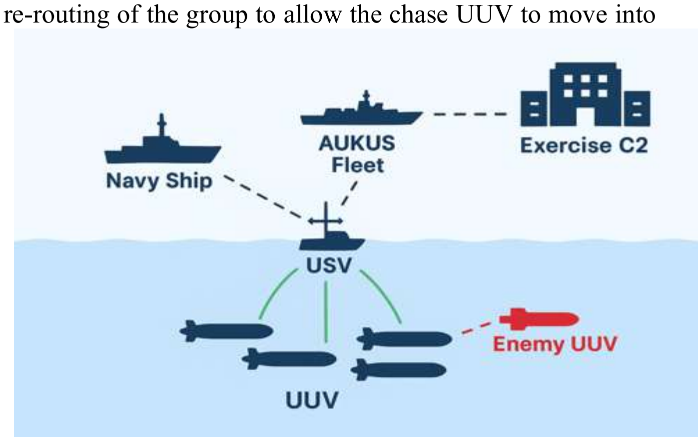


**Figure 2 – UUV Concept of Operations**

The patrols are set for an approximate 30-minute mission. This would allow a reasonable bounding box for operations, recovery and reset for the next mission. The variation of the starting positions and position / velocity vector of the enemy provides a robust testing environment.

The UUV for this exercise is provided by VATN technologies. It is a 6 inch by 6-foot UUV with a suite of surface and underwater sensors. For the purpose of the concept of operations, the platform speeds are set to 3-4 kts with a conservative turn radius. They have onboard FLS and acoustic communications devices to find the enemy and formation vehicles. The acoustic communications is a low cost efficient method to have the vehicles communicate with each other and with the surface. For the simulation setup, it is assumed that the sonar can work to its specified maximum 50-meter range and the underwater communications devices can run to their maximum of 1000 meters. Typically, the platforms are started/stopped by either network commands or timed starts/stops.

## CREATION OF UNMANNED BEHAVIORS

Recent advances in artificial intelligence and machine learning (AI/ML) are transforming the development of collaborative behaviors for unmanned underwater vehicles (UUVs). Simulation-driven multi-agent reinforcement learning (MARL) frameworks are being used to train UUV teams to perform cooperative missions such as formation control, target tracking, encirclement, and area coverage in communication-constrained and noisy underwater environments. Platforms like MarineGym (Chu et al., 2025) offer GPU-accelerated, high-fidelity physics environments that allow multiple UUV models to be trained simultaneously with algorithms such as Proximal Policy Optimization (PPO) and Multi-Agent PPO (MAPPO). Recent research has shown that incorporating domain knowledge, such as artificial potential field (APF) guidance or neural control priors, significantly accelerates learning in complex tasks like cooperative hunting and obstacle avoidance (Zhang et al., 2020; Praczyk & Szymak, 2023). These efforts are further strengthened by simulation platforms that allow realistic sensor modeling, such as sonar, ocean current disturbance, and localization noise, improving the transferability of learned policies to physical vehicles. (Drewes 2009)


**Figure 3 – UUV for Collaborative Operations**

The state-of-the-art also includes work on hierarchical and Transformer-based policies that enable UUV swarms to coordinate complex tasks in dynamic environments with limited acoustic communication. For example, Gallici et al. (2025) used a scalable MARL approach based on a Transformer policy architecture (TransfMAPPO) to train swarms of underwater vehicles for collaborative tracking of moving targets under acoustic channel constraints. Other studies have combined deep reinforcement learning (e.g., TD3, DDPG) with simulation-derived feedback, such as APF trajectories, to support swarm behaviors even when environment observability is weak (MDPI, 2021). These advances mark a critical shift from manually programmed behaviors to data-driven autonomous systems capable of adapting in real time to complex underwater environments, paving the way for operational deployment of intelligent UUV swarms in mine countermeasures, ISR, and collaborative manipulation missions.In addition, studies involving AI/ML with AFSIM demonstrates how reinforcement learning (RL) agents can be trained within a mission-level simulation to perform tasks such as leader-following and object sensing (RAND). Their approach connected AFSIM to Python-based ML frameworks, allowing AI agents to learn behaviors in a simulated operational environment. However, the study also highlighted significant limitations: AFSIM was not originally designed for dynamic, in-simulation learning, making real-time training and adaptive decision-making difficult without heavy external control. Additionally, the simulation's discrete event-driven nature and simplified physics can limit the fidelity of environmental feedback critical for certain types of machine learning algorithms. These constraints necessitate careful tailoring of the AI models and often require significant engineering overhead to align AFSIM's architecture with modern AI/ML workflows.

### Simulation Environment for Baseline Behaviors Using AI/ML

Our approach was to connect AFSIM to an AI/ML training system to adapt generic UUV behaviors and responses into a series of group behaviors. The goal is to integrate these series into UUVs. These behaviors make up specific parts of missions. They could be considered plays in a playbook during a football game. Each team member (UUV) knows its role and the role of everyone in the team. The purpose is to minimize the run time communications between the team members once a play "has been called". Just like a wide receiver runs down the field and perceives there is no football coming, the UUVs can perceive and adapt to support the overall play in progress without needing additional coordination.

The core simulation was taken from AFSIM and underwater vehicle motion. Since the VATN S6 vehicles are not part of the core AFSIM baseline, a generic set of adaptations were made concerning their behaviors. This included velocities between 1-4 m/s, and depth ranges between 5-20 meters. A series of 20,000 simulations were created to train the AI/ML models. Since this was offline learning to the simulation environment, a series of data logs were created concerning the output state of the individual UUVs under test. For the initial tests, four UUVs were created in various formations to detect and track a single inbound adversary UUV.

The simulation output tracked several onboard parameters and tracking information. This included relative UUV positions from each other and the adversary when detected, individual unit speed, heading, battery life, and detection data of the adversary including range, bearing, track speed, and heading. Real world items such as current differentiation, heading inaccuracies, and position error were not included in the baseline configurations. This data was fed into a series of test models which attempted to optimize the formation integrity, and minimize inter-UUV collisions, energy and maneuver cost for tracking the UUV while maximizing the coverage and surveillance range of the remaining UUVs. This fed a physics-based feature dataset. This included the individual UUV distance to intercept, angle of intercept and the reformation cost of the remaining UUVs to reform once one departs the formation. All these parameters are scaled using min-max normalization and combined into a weighted cost function as shown in Figure 4 – Total cost calculation. The UUV with the lowest total cost is selected to pursue the target.

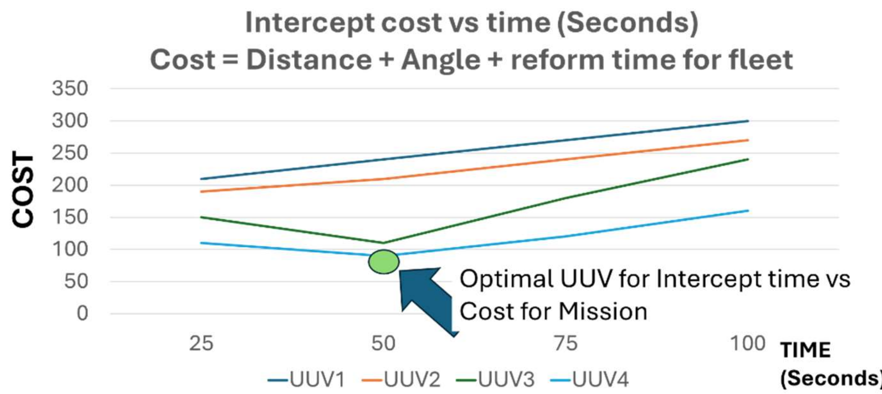


**Figure 4 – Total cost calculation**

Given the complexity of the mathematical equations needed to extend the physics-based approach to further domains like probabilistic data, such as battery life and red team events, a neural network was employed to illustrate how one might approximate the relationship between the input data and the optimal UUV behavior. This model was composed of an input layer sized to ingest both datasets discussed above. A series of hidden layer tests were run. From one to six neural layers were tested and the difference in performance above four seemed to indicate minimal differences. The four hidden layers decreased in size from 128 to 32 neurons (Figure 5). Each hidden layer used REctified Linear Unit ReLU as its activation function. The output layer was 4 neurons that used softmax as its activation function. The loss function selected for the training exercises was categorical cross entropy which was well suited for this multi-class classification problem.

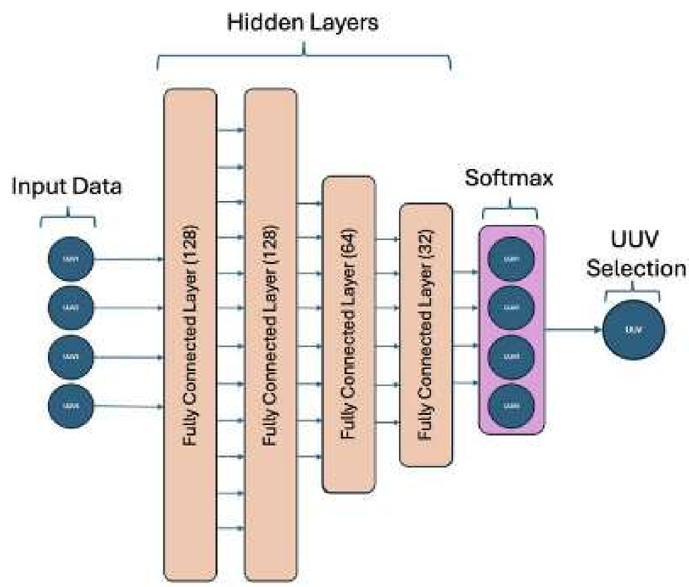


**Figure 5 – Neural Network Design**

Battery life was incorporated to illustrate how this approach could be extended to include additional features and considerations. This included randomly selecting each UUV's battery life between 0% and 100%, and adding a probabilistic element to the initial cost function. This produced two sets of labels. The first set of labels on the left shown in Figure 6, shows the optimal UUV selected without considering battery life. The set of labels on the right considers battery life. It is apparent in these distributions that the data is heavily skewed towards selecting UUV1. This is due to the adversary positioning and the cost function in other bots leaving the network and reforming the grid. One would expect the end UUVs (one and four) in the formation to be higher.

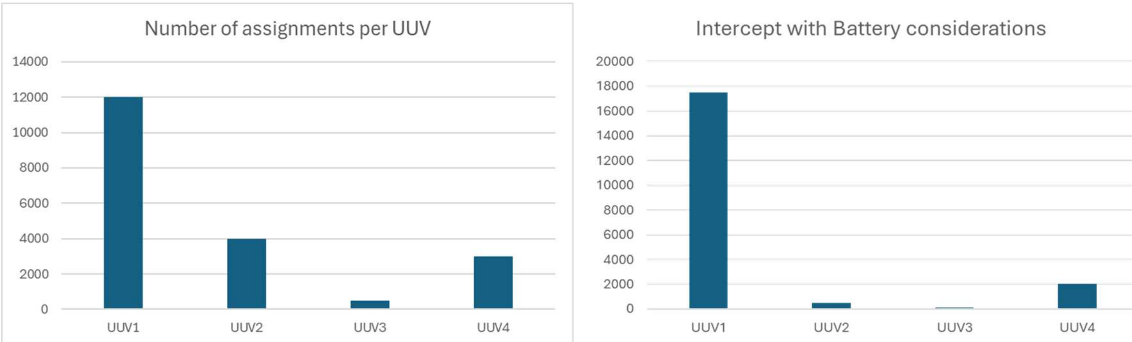


Figure 6 – Class Distribution

**Training and Simulation results for AI/ML Model**

A training exercise was performed for both of the discussed datasets. The model was trained on the entire 20,074 data samples. The data was split using 80% for training and the remaining 20% for validation purposes. The model was trained for 2,000 epochs. The Adam optimizer and a learning rate of 1E-4 was selected through a series of testing where these respective values yielded the best results. Model weights were restored to the epoch that produced the best validation accuracy.

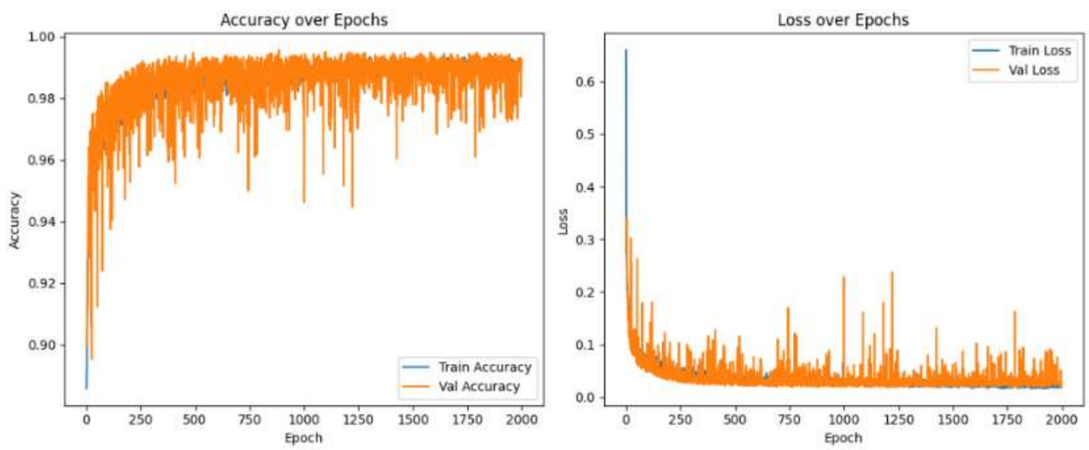


**Figure 7 – Training accuracy**

Figure 7 shows the training and validation accuracy, as well as the training and validation losses for the training exercise with battery life considered. The best validation accuracy found was 99.55% for this exercise. Figure 4 shows the same performance metrics for the training exercise that incorporates battery life. The best validation accuracy found was 97.73%. The training accuracy is import from the theoretical perspective to show the model is accurate based on the training data from the simulation environment.

The experiment demonstrates that team based UUV operations can be significantly enhanced by intelligent task allocation mechanisms. By combining rule-based physics models with flexible neural network architectures, the decision model can both scale and adapt to mission complexity as additional non-linear data becomes available. The neural networks quickly allowed for new factors such as underwater acoustic simulation latency to be added with minimal development overhead. The models auto generated were able to match the hand coded physics models in accuracy and efficiency, resulting in faster model turn around than hand coding behaviors.

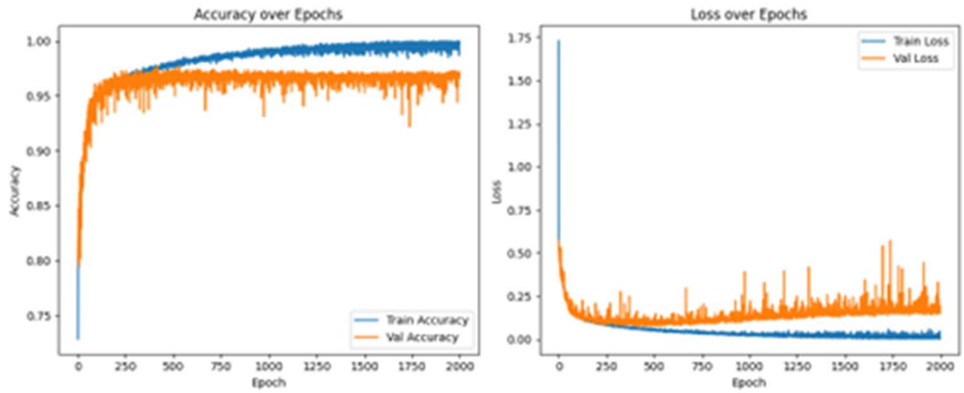


**Figure 8 – Validation Accuracy**

**Verification & Validation (V&V) of Behaviors**

Once the behaviors were tested in an AI/ML test environment, they were shifted to testing by the UUV simulated / live and surrogate testing. The overall metrics of the detection and tracking for the program were set at the 70% level to account for times when the vehicle might not be detected or interference with other underwater objects results in less than a hundred percent testing environment. To test the collaborative behaviors individually and as a group the tests were setup in several environments.

Since the AFSIM test environment was not used with the UUVs and these UUV configurations are not yet part of the AFSIM baseline, the simulated test environment was done differently. The feedback loop was connected to the vehicle provider's (VATN's) digital twin of the vehicle behaviors. This was done through a basic TCP/IP network socket connection to allow simulated entities to be tested. The challenge is that of accuracy of the acoustic communications and the FLS sensor algorithms. Each individual and group behaviors was tested within the *perfect* simulation environment of the digital twin before testing on live platforms. This provided a validation that the code should work as expected assuming the live conditions are similar. However, for the testing timeframe, the FLS sensor could only create generic output. There was no way to stimulate or replicate the sensor data. So, offline from the digital twin were data logs that could be replied and sent to the playbook to examine the V&V accuracy of the behaviors.

In addition to the FLS limitations, the acoustic communications are set as basic messages that always reach their destination given a transmission. Transmitting acoustic data underwater is significantly slower than electromagnetic communication in air due to the physical properties of water. Sound travels underwater at roughly 1,500 meters per second, which is much slower than the speed of light. This causes high latency, especially over long distances. Additionally, acoustic signals are affected by factors like multipath propagation, absorption, ambient noise, and variable sound speed profiles due to temperature, salinity, and pressure. These limitations make high-speed or real-time communication difficult. Testing such transmissions realistically in a basic simulation environment is challenging, as accurate modeling of underwater acoustics requires complex physics and environmental parameters. However, simplified simulations using fixed propagation delays and packet loss models can offer a basic approximation for early-stage testing or protocol design.

To assist with some of the testing, surrogate platforms and sensors were used to verify the behaviors. A full set of leader/follower test behaviors were done on small ground robots using the Robot Operating System (ROS) and Light Detection and Ranging (LIDAR) to emulate the tracking of sonar objects. Using ROS to set up a surrogate test of sonar detection and tracking for an UUV using a ground-based UGV with a LiDAR system offers several practical advantages. ROS provides a flexible, modular framework that supports rapid prototyping and integration of sensing, control, and communication systems, making it ideal for simulating UUV behavior in a more accessible terrestrial environment. By substituting sonar with LiDAR and water dynamics with ground-based navigation, the team tested detection, tracking, and collaborative behavioral plays without the logistical and environmental challenges of underwater testing. This limited test environment allows a more robust testing of the top-level play behaviors before the sonar data was available.

Once the simulated behaviors and controls were tested using the surrogate and the digital twin provided by VATN, the in-water testing began. This reduced the testing time using live underwater platforms by over 50%. This was measured as the developer time spent in simulation hours vs the actual in water testing time. The in-water testing time was measured by the time to run a test plus the setup and hardware repositioning time which was much larger than the actual testing time.

### Feedback Loop for Behaviors

Traditionally, the feedback loop would be back to the AI/ML baseline as well as updating the core simulation models. However, since this research effort doesn't include updating the official AFSIM baseline, the feedback loop is done with the digital twin models and the AI/ML modeling behaviors. It is possible to update and/or create a traditional constructive simulation extension for the group behaviors. However, the primary goal was to update the behavioral playbook which is stored in the AI/ML models for the live vehicles.

## LIVE TESTING OF PLATFORMS

### Sensor Coverage

The primary sensor on the UUVs is the FLS system. It is a 720Khz sonar system with a specified range of 50 meters. Early testing indicated that the full range is available for detection and potential identification of UUV targets. Figure 9 shows both the profile of the sensor and a representative return from the sonar system. The target environment for the testing is offshore in 15-30 meters of water. This will allow the sensor to perform at its full range without surface or seafloor interference. The FLS data is processed via the onboard vehicle processor and summary data concerning targets is passed to the playbook system. This reduces the communications between the sonar collection algorithms and the decision points in the system.


**Figure 9 – Gemini FLS**

The system is also equipped with a Doppler Velocity Log (DVL). The DVL is a critical navigation sensor for the system providing precise velocity measurements relative to the seafloor by using acoustic Doppler shifts. In shallow water environments such as 15–30 meters depth, a DVL is especially valuable because it can reliably maintain bottom lock—essential for accurate dead reckoning when GPS is unavailable underwater. For experimental goals involving tracking and positioning in this depth range, the DVL offers high-resolution velocity data that helps correct drift from inertial sensors and enables stable, consistent path estimation. This enhances the fidelity of behavioral experiments, such as formation control or waypoint navigation, by ensuring the UUV's estimated position remains accurate enough for meaningful evaluation of collaborative or autonomous strategies. This allows the GPS like positioning underwater to better verify the positions relative to each other and the enemy positions.

### Simulator and live testing

A combination of testing was done throughout the baseline configuration. The AI/ML tools ran their own V&V process to examine the data differences and behaviors from the AFSIM baseline for process improvement. During early development, additional ROS based simulations and hardware emulators were created to test some of the multi-bot communications and LIDAR functionality. This reduced risk in testing with the hardware provider's simulation as it matured capabilities after some of the testing was complete. Surrogate hardware and software was used to test early behaviors and motion models. This mini LVC test environment reduced the risk and time needed to test in the live target platform in the water. The live in water testing is the most expensive part of the testing phase. It is needed since it is the only environment to provide full sonar data from the communications, FLS, DVL and GPS errors along with the collaborative behaviors. By feeding this back to the various simulations, they can be more representative of the operational environment. This was done in a limited fashion due to the testing timeframes involved.

Communications between UUVs underwater and any C2 environment above water is the most challenging portion of the testing. The target detection distances (30-50 meters) and fleet horizontal distances (5-20 meters) made any communications methods beyond acoustic difficult. So a series of tests and digital twin models were created to ensure that the communications could be adapted to ensure optimal communications. This included updating the fleet on the position and mission play underway balanced against the loss of data if the UUVs step on each other since they run broadcast messages. Even though acoustic communications can run in the hundreds of meters, the bandwidth is a function of vehicle velocity, distance between transmit/receive and environmental conditions (temperature, salinity, surface/seafloor distance). A representative model was created to inform the communications model on how much could be shared and when. This resulted in a variable optimal communications model based on the testing environments to update as quickly as possible while minimizing the loss.

**Results**

The systems were tested in the live water environment. This allowed the changes to the core algorithms based on a noisy water environment. This included surface, seafloor and debris interference to the sonar systems. This resulted in two key elements. The sonar data required additional filtering to find and assist in the tracking of the adversary. The motion and water resulted in less-than-optimal communications between the active teammates. This resulted in time delays in communicating the play between the UUVs and executing the plays. The delay caused the team plays to sometimes break where one UUV would detect and coordinate another that will track it. However, by the time the tracking bot was made aware of its role, the adversary was not in range of the receiver. The original plays were modified to reduce communications and to do more prediction to increase the intercept probability.

The team took data logs of each run including communications, location/orientation and raw sonar data for post processing and additional AI/ML analysis of what was detect when and how well it was tracked. This allowed for additional simulation testing based on live data to update the plays and core UUV behavior algorithms. The live data concerning position and adversaries was connected to the surface and then to Starlink to integrate into the RAAIT data collection and command/control environment for a full situational awareness.

**Summary and next steps**

Based on the results, the simulations of the systems, and the plays/behaviors were updated. The only missing connection was updating the AFSIM models, but since the UUV’s aren’t part of the baseline, this was saved for a potential future update. The simulation testing V&V and feedback loops allowed the team to be able to perform hardware/software integration and testing in under six weeks. The learned behaviors were able to be executed on the live UUVs taking into consideration the limitations of the communications and positioning accuracy underwater of the UUVs. This allowed a swarm of UUVs to be trained to behave offline, then to detect and react to adversaries while minimizing movements between the fleet to track and watch for the next adversary. This mature playbook works well for a four group of UUVs running a counter UUV solution.

For next steps, the team wants to work additional plays, such as he who finds it follows it, or using heterogenous platforms to perform different roles instead of the existing homogenous teams. The live feedback to the core learning model could be enhanced by connecting it to the hardware vendor’s simulations. It might make more sense to integrate the simulation into an environment such as the Robotic Operating System (ROS) or a commercial game engine like Unreal that provides easier plug in for hardware in the loop testing, proprietary simulation models or live updates. The AI/ML updates provided a good baseline to start testing. A full feedback loop based on changes needed, sensor and communication limitations would potentially provide new plays to add to the playbook.

**REFERENCES**

Drewes, P., Jameson, S., (2009), "Unmanned Surface Systems Collaborative Experimentation," in the Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan, May 12-17, 2009

RAND study and analysis. " Understanding the Limits of Artificial Intelligence for Warfighters", retrieved May 1 2025 from https://www.rand.org/content/dam/rand/pubs/research_reports/RRA1700/RRA1722-5/RAND_RRA1722-5.pdf

Chu, S., Huang, Z., Lin, M., Li, D., Carlucho, I., Petillot, Y. R., & Yang, C. (2025), "MarineGym: A high-performance reinforcement learning platform for underwater robotics", retrieved June 1 from https://arxiv.org/pdf/2503.09203?

Gallici, M., Masmitja, I., & Martín, M. (2025), "Scaling multi-agent reinforcement learning for underwater acoustic tracking via autonomous vehicles" , retrieved June 1 from https://www.researchgate.net/publication/391706557_Scaling_Multi_Agent_Reinforcement_Learning_for_Underwater_Acoustic_Tracking_via_Autonomous_Vehicles

Zhang, Q., Lin, J., Sha, Q., He, B., & Li, G. (2020). "Deep interactive reinforcement learning for path following of autonomous underwater vehicle" . Retrieved June 1 from https://www.researchgate.net/publication/338547093_Deep_Interactive_Reinforcement_Learning_for_Path_Following_of_Autonomous_Underwater_Vehicle

Praczyk, T., & Szymak, P. (2023). "Neural swarm control algorithm for underwater vehicles", WSEAS Transactions on Systems and Control, 18, 300–306.

MDPI. (2021). "A cooperative hunting method for multi-AUV swarm in underwater weak information environment with obstacles", Journal of Marine Science and Engineering, 10*(9), 1266.