Free Jazz on the Battlefield

How GhostPlay's AI Approach Enhances Air Defense

Heiko Borchert*
Co-Director Defense AI Observatory
Helmut Schmidt University
University of the German Armed Forces
Hamburg, Germany
*hb@defenseai.eu

Armin Brandstetter*
Hensoldt Sensor Systems
Ulm, Germany
*armin.brandstetter@hensoldt.net

Christian Brandlhuber*
21strategies
Munich, Germany
*christian.brandlhuber@21strategies.com

Gary S. Schaal*
Co-Director Defense AI Observatory
Helmut Schmidt University
University of the German Armed Forces
Hamburg, Germany
gschaal@hsu-hh.de

***Abstract* – Current conflicts underline the importance of Integrated Air Defense Systems (IADS) to keep aggressor air power at distance and ensure allied freedom of maneuver. But what happens if aggressors saturate, deceive, and neutralize allied air defense with hundreds of unmanned aerial assets in conventional attrition attacks or apply hitherto unknown tactics potentially enhanced by artificial intelligence (AI)? That's the question GhostPlay addresses by developing defense decision algorithms (= Play) to support tactical military decision-making against aggressors that operate at different levels of ambition, excel at leveraging unknown and emerging tactics, and strive to exploit operational tempo to their benefit. GhostPlay uses a synthetic simulation environment (= Ghost) to assess if and to what extent AI-enhanced solutions – operating in stand-alone or federated systems – can be used to accelerate operational tempo, enhance tactical level performance, and step-up efforts to anticipate future adversarial behavior. Against the background of a growing body of literature on defense innovation, the paper discusses GhostPlay's goal to develop context and consequence-aware AI systems that exploit novel tactics to ensure and scale IADS-based protection. The paper sheds light on GhostPlay's conceptual and technical setup, summarizes initial simulation-based findings and outlines future development options.**

Keyword – Defense Artificial Intelligence, emergent behavior, multi-agent systems, swarm logic, tactical versatility.

NOMENCLATURE

A2AD	Anti-Access/Area Denial
AAA	Anti-Aircraft Artillery
AD	Air Defense
AI	Artificial Intelligence
C2	Command and Control
C4	Command, Control, Computers, Communications
DARPA	Defense Advanced Research Projects Agency
DecPOMDP	Decentralize Partially Observable Markov Decision Process
EmCon	Emission Control
EW	Electronic Warfare
FlkPz	Flakpanzer
HARM	High-Speed Anti-Radiation Missile
HVA	High Value Asset
IADS	Integrated Air Defense
JTFS	Joint Tactical Fire Support
MDP	Markov Decision Process
OODA	Orient, Observe, Decide, Act
POMDP	Partially Observable Markov Decision Process
RAP	Recognized Air Picture
RL	Reinforcement Learning
ROE	Rules of Engagement
SHORAD	Short-Range Air Defense
UAV	Unmanned Aerial Vehicle
UCAV	Unmanned Combat Aerial Vehicle
VBE	Value-Based Engineering
VSHORAD	Very Short-Range Air Defense

I. If Innovation Is the Solution, What Is the Problem?

"Military Innovation" has become a hot buzzword among NATO and EU members. Two forces shape this current defense innovation discourse. First, increasingly assertive adversarial military capabilities underline the need for allied defense innovation to keep the upper hand vis-à-vis strategic competitors [1]. Second, the defense innovation discourse emphasizes the important role of emerging technologies like artificial intelligence (AI), autonomous and robotic systems, space, and quantum technologies to name but a few examples [2]. In most cases, commercial entities that are not yet part of the defense ecosystem are frontrunners in developing and applying these technologies. This increases the need to integrate new players, technologies, and underpinning capacities into the defense industrial and technology base.

Although in fashion, defense innovation is notoriously difficult to define [3]. Few capstone documents describe precisely what type of innovation armed forces are expected to deliver and what needs to change to accomplish the respective tasks. Based on [4] we contend that defense innovation describes conceptual/cultural, organizational, and technological novelties that change how armed forces prepare for and conduct the application of military power. In so doing, armed forces build on past operational experience and requirements.

Against the background of these three vectors, GhostPlay's innovation understanding is two-fold. First, GhostPlay addresses a pressing gap as Suppression of Enemy Air Defense (SEAD) capabilities have atrophied in most EU/NATO nations since the end of the Cold War. We explore to what extent AI-based solutions can augment swarms of unmanned aerial vehicles (UAV) to conduct SEAD missions. Second, GhostPlay does not look at new technologies to augment existing technologies. Rather we look at ways in which the use of new technology triggers novel battlefield behavior at the tactical level. With these two aspects in mind, GhostPlay models novel AI-based solutions for air defense (AD) and aggressor swarms that learn how to outperform each other. The first project phase, which we discuss in this paper, focuses on the defender.

To model and learn superior tactical AD behavior that withstands and counters UAV swarms, we consider two main aspects. First, in most recent conflicts UAVs gained the upper hand against AD as AD solutions have been brittle [4]. Brittleness results from a lack of proper integration of all relevant sensors and effectors to create a powerful AD federation. Integration, in turn, requires coordination. This is where the second element kicks in. GhostPlay focuses on novel approaches that increase tactical AD versatility to fend-off aggressors. In so doing, GhostPlay breaks new ground by exploring options to develop federated AD webs that coordinate single entities like sensors and effectors through emergent behavior without the help of central and hierarchical command and control (C2) solutions. As we explain in section II.B, GhostPlay bakes the C2 capability into every element of the AD web rather than delegating C2 to a dedicated system, that adversaries can target and attack. This approach makes the AD web much more fluid, agile, and resilient in responding to threats and mission requirements.

Superior tactical versatility augments military freedom of action. To this purpose GhostPlay seeks to leverage the principles of war that guide and inform how military power is applied [5]. Among other things, GhostPlay strives for

- economy of effort by optimizing the use of effectors in time and place as well as with respect to how force is organized to achieve optimal effects under any given conditions;
- surprise by using emergent behavior in a way that produces tactical behavior not yet witnessed by aggressors;
- initiative by anticipating future adversarial moves with the goal to preemptively position allied force to engage adversaries.

In sum, GhostPlay contributes to defense innovation by developing technology that enables novel battlefield behavior to enhance tactical versatility, first, for air defenders and, at later stages, also for UAV swarms performing SEAD missions. In this regard, GhostPlay's innovation is like free jazz as it improvises, responds to external stimuli, is dynamic, and integrates whatever asset is available to accomplish the AD mission by leveraging a new generation of coordination mechanisms that are context and consequence aware.

II. GhostPlay's Novelty: Free Jazz vs. Central Coordination

While GhostPlay strives to create innovation in terms of tangible advantages and capability improvements for future AD concepts, the project's underlying technology contributes to one of the most challenging topics in contemporary AI research, the ability to learn tactical behavior in cooperation with other machines and/or humans. This entails three capabilities. First, the capability to properly assess a situation and anticipate adversarial behavior. Second, the capability to learn how to orchestrate and organize a system's action to achieve objectives across time-extended scenarios and in response to enemy action. This also includes the ability to assess, how the relevant environment may respond to the defense system's actions. Third, the capability to motivate a system to learn on its own when and how to cooperate to solve complex tasks with partners. These capabilities underpin future solutions striving for technical autonomy in machine-to-machine and machine-to-human interaction.

Right now, the idea that Deep Reinforcement Learning solutions like AlphaGo, Alpha Star or Open AI have super-human capabilities creates quite a hype. But these systems play computer games in a well-known and completely stable environment. Military solutions, by contrast, operate in a non-stationary real-world environment, where unforeseen incidents occur. Moreover, commensurate with adversarial intentions and capabilities, the rules of the military game can change quite quickly.

This is the environment in which GhostPlay is supposed to operate. Integrated Air Defense Solutions (IADS) adopt a layered approach (Figure 1). Sensor and effector reach is the discriminator that helps setting up Very Short Range Air Defense (VSHORAD), Short Range Air Defense (SHORAD), Medium Range Air Defense (MRAD) and long-range defensive ground-based "domes."

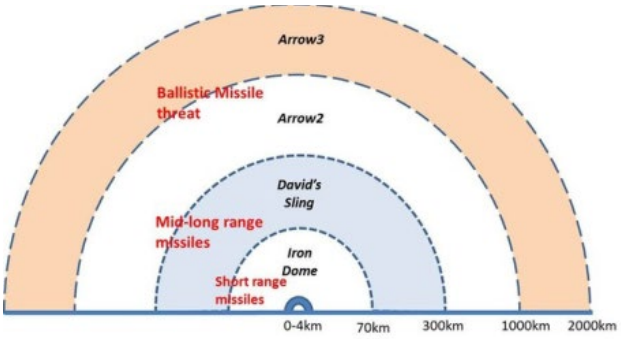

FIGURE 1: LAYERED APPROACH TO AIR DEFENSE. SOURCE [10].

Today, specific systems integrate different sensors and effectors for each "dome." Each system is developed in isolation. The governing principle to achieve integration is hierarchical and centralistic: Central command and control (C2) runs each system, which report into a hierarchic structure with nodes that coordinates multiple launchers. Ultimately, a high-level C2 or C4 system (Command, Control, Communications and Computers) integrates and coordinates all elements. Although tried and tested, this set up remains static, is quite brittle and often leads to unsatisfactory sensor-to-effector latency.

Contemporary state-of-the art solutions may use AI to improve individual processing steps in the observe-orient-decide-act cycle (OODA). GhostPlay goes beyond the automation of individual steps in the OODA workflow. Rather, GhostPlay policies establish a fine granular and forward-looking stochastic optimal control regime, while substantially accelerating decision-making and reducing sensor-to-effector latency.

The GhostPlay architecture achieves this by mediating multiple concurrent control processes ("agent"), each implementing a specialized control strategy ("policy"), for example, to effectively control a physical system, like sensors and effectors, or to determine a certain action plan. GhostPlay's agents are not centrally coordinated. Rather they use common behavioral conventions ("rules of encounter") to ensure mediation and information exchange while training to achieve a common objective in tandem with partners (multi-agent learning). In contrast to existing AD solutions, GhostPlay has no pipelined data-fusion process on the platform; rather a combined situational picture emerges over time via policies that have learned how to cooperate. This specific design choice has been motivated by the ambition to explore if defense solutions can be developed as emergent systems.

A. Tactical AI: Basic Building Blocks

1) State of the art defense AI

Currently, state of the art defense AI focuses on introducing AI techniques or AI-based components to support individual OODA stages. For example, current applications detect and classify objects in aerial reconnaissance pictures or classify characteristics in the electromagnetic spectrum to infer a potential emitter. In so doing, the OODA loop implements (linear) pipeline processing. Thus, it is sufficient that deployed AI systems only implement a single-step decision. For example, an optical sensor acquires an image or a single video frame, if the sensor streams video sequences (observe); this image is sent to an AI component that tries to detect and classify objects (orient); detected objects will be sent further down the pipeline for identification, and finally arrive at a decision to complete the cycle. The AI component can be interpreted as a Bayesian classifier, which determines the probability that a certain (known) object is present based on the input image and the parameters the AI component has learnt, in most cases weights of a neural network. Currently, the AI classifier thus implements a one-step input-to-output mapping.

2) Improve classification and identification in non-benign environments and fog-of-war

In most real-world applications, in which environmental effects cannot be controlled,[1] or – even worse – if the sensor must operate in a non-benign environment, better results can be obtained, if classification is integrated in a context-aware sensor control strategy (policy): The optical sensor acquires an image, which shows multiple objects. As these objects are far away from the observer they only appear as unspecific pixels on the image. A state-of-the-art AI classifier approach would have to classify each picture and would most simply ignore these target pixels.[2] By contrast, a policy-based sensor management system has different options: Based on the input image, the policy may decide to zoom-in on those pixels which are presumed to represent potential targets or threats that require additional sensor input for classification. The system might even decide to illuminate the position of interest to obtain better information, for example, if clouds obscure the respective objects. This approach would enable the system to provide stable classifications early on and long before the state-of-the-art approach would even be able to recognize the object.

It is very important to note the difference between these two approaches: Interlacing sensor management with classification and identification requires the system not only to learn input-to-output mapping as current AI deployments do. Rather the system needs to learn courses of actions to understand how to achieve good classification results as efficiently as possible. To learn good action sequences, the system must specifically learn to understand the instant effects of an action as well as possible long-term consequences. From a mathematical viewpoint this is no longer a Bayesian classification task, but requires solutions based on the mathematical framework of a Markov Decision Process (MDP).[3]

3) Learn important trade-offs to maximize effectiveness and minimize own-ship exposure/risk

In addition to understanding the instant impact of a specific action, learning good policies also implies that the system learns how to make important trade-offs. Zooming-in on a position, for example, reduces the observation window and might lead to a situation where the system "gets stuck" on pedantically classifying one object, while not recognizing that a fast-moving threat is heading towards the system outside of its observation window. The system also needs to carefully

[1] Industrial environments typically try to create controllable environmental conditions such as lighting. For example, objects that come down a conveyer are always sensed in the same light to limit/exclude negative external effects on conveyor belt transportation.

[2] Note, that this is not the result of a bad classifier but rather results from the fact that it is not possible to classify the targets with the given sensor input.

[3] As the system perceives its environment through sensors and thus only has imperfect perception, we cannot assume to correctly observe the true states of the MDP. Rather we estimate and partially reconstruct the true state from the sequence of observations made so far, which further complicates the task into a Partially Observable Markov Decision Process (POMDP).

balance short-term success with long-term consequences: For example, deciding to illuminate a target with an active radar sensor may satisfy a short term information need but can put the observer at risk as the radar signal emitter may be detected by the target, which in turn may fire a high-speed anti-radiation missile (HARM) to destroy the observer.[4][5] If the benefits of using an active sensor are bigger than the risks very much depends on the situational context. Moreover, the decision to use the active sensor directly affects how the situation will evolve. Systems that master this complexity are context-sensitive and consequence aware and constitute so called 3rd wave AI systems according to the US Defense Advanced Research Projects Agency (DARPA).[6]

Most often Deep Reinforcement Learning is used to train these systems, but this creates technical challenges. Take AI concepts for classification as an example. There is a tutorial input (e.g., tagged example) for each decision made by the system after each round of classification. But there is no immediate feedback to the system, which allows the system to understand whether the respective decision influenced the scenario in a positive or negative way.[7] Although the system needs to maximize the long-term reward intake, the missing link suggests that positive or negative decision outcomes will only be known at the end of the scenario. This, however, can involve several thousand decision steps into the future, which means that the system may get out of sync with the proper function it is expected to accomplish.

4) Create good initial policies without large databases

There are concepts in reinforcement learning training protocols that collect traces through scenarios to aggregate so called state-values V(s) or state-action values Q(s,a). These concepts show, if – on average – selecting action a when in state s has been good or bad. Based on these results neural network structures are trained to represent the respective value functions. However, using only these concepts in practice may lead to (very) sub-optimal policies, because the positive or negative outcome of selecting action a in state s not only depends on the current state, but also on the policy that guides future actions. For example, a tracking radar is switched on to illuminate a target and gain accurate position and movement estimates for engagement. This decision may be positive if the threat is successfully intercepted. But it can also be negative in case of failure as the system has exposed its position and created an opportunity for adversarial attack. Moreover, using less precise passive sensors to preassign targets while reducing exposure time would have improved tactics. To gradually converge to good policies, reinforcement learning systems need to strike a balance between exploiting past behavior and exploring new behavior that could deliver novel policies.

With scenarios spanning thousands of decision steps, finding good policies becomes combinatorically prohibitive. This creates specific issues during early training phases. In almost all cases, in which applications have been said to have "superhuman" decision-making capability, initial policies used to start reinforcement improvements have been developed with supervised learning. Supervised learning, in turn, was possible because large databases of expert level policies were available to create tutorial input. But the military application area addressed by GhostPlay, lacks the respective databases. That's why GhostPlay needed to find a way to create initial good policies without databases.

Today, GhostPlay implements a novel "search-in-policy-space technique" to achieve this objective. We decided to initially model an air defense platform, which has multiple onboard sensors and one effector. Each sensor and effector has its own policy, which learns how to optimally use the sensor's specific characteristics. Data is exchanged via a central on-platform long-term memory structure, from which all policies can read and to which all policies can write. Cooperation amongst the policies is mediated by a stigmergic signal. As expected, the resulting platform behavior is rather complex and adapts to fine nuances of an emerging scenario.

B. Emergence: Cooperative behavior paves the ground for technical autonomy

Success in joint problem solving very much depends on the way in which perception and interaction with other agents in the team are modeled. As discussed above, a classical AD setup collects and propagates information via different sensors to a central C2 node, where information is aggregated, fused and appropriate courses of action are calculated. Then orders and instructions are flowing down the chain of command to individual effector systems, in our case the anti-aircraft artillery (AAA) platform. This approach is tried and tested but also raises several issues:

a) Network centricity: The process heavily relies on transmitting data through the network to and from C2 nodes, which largely coordinate individual platforms, unless they operate in self-defense mode. What if communication is distrupted and bandwidth is limited? Are there other ways to reorganize local entities for effective cooperation if communication breaks down?

b) Sensor to effector latency: Propagating information through networks that require C2 nodes for data fusion generates sensor-to-effector latencies. Latency, in turn, can put individual AD platforms at risk if incoming threats are detected too late.

c) Single point of failure: C2 nodes may constitute a single point of failure. If the opponent manages to detect and take out the C2 node, the whole AD network becomes ineffective or at least massively degraded.

d) Reconfiguration: Even if the C2 node is not affected, loss of individual sensors or effectors in the network may require a reorganization of the compound. Currently this requires replanning, which again results in latencies. Looking at attrition scenarios, we assume that the ability of some network elements to automatically regroup could substantially improve overall resilience and effectiveness.

[4] Balancing short-term reward intake with long-term objectives is part of the "temporal credit assignment problem."

[5] The actual implementation requires skillful engineering of what constitutes an "action:" The system must learn how to use radar functions in a sensible way, for example, by allowing a tracker to initialize and maintain a track with reasonable accuracy. We are currently working with macro-actions, which provide complete implementations for certain tasks. Moreover, we are experimenting with a combination of "track-before-detect" and "attention-based tracking" to analyze if these trained, model-free variants would enable faster effector engagement and better self-protection in high treat scenarios.

[6] https://www.darpa.mil/about-us/darpa-perspective-on-ai.

[7] AI classifiers typically use tutorial training. In this case, the immediate tutorial input is used to form an error signal, which is backpropagated into the classifier to adjust those parameters, which had the highest contribution to the error.

e) *Ad-hoc support:* Attackers commonly exploit the "relative strength principle." This means that attackers will try to concentrate force at a specific and narrow point of the defender to temporarily overwhelm it. Even if the defender massively steps up its efforts, it is almost impossible to avoid that forces at the point of attack quickly run out of ammunition, while the larger part of the defense infrastructure is almost unaffected. We speculate, that a system, which is able to locally reorganize, can provide ad-hoc support to the very forces under heavy attack and reinforce them quicker.

f) *Economy of effort:* Multiple systems of an IADS cover the same airspace. In practice each of these systems has its own C2 component. These C2 components need to decide or negotiate which effectors to deploy, such that economy of effort is preserved. This decision is highly context sensitive.[8] We assume that a system that properly understands this context will be able to make more effective effector choices commensurate with the threat.

To explore these hypotheses, our objective was to experiment with a setup, which does not have a C2 component at all. Rather our system is composed of individual AD platforms that learn how to cooperate and find an effective and emergent defense response against any incoming threat.[9] In essence, we strive to learn policies, which motivate other agents in the same team to cooperate. To do so, we model joint behavior amongst our AAA platforms as a Decentralized POMDP (DecPOMDP).

The general idea is to develop a "theory of mind" among agents, i.e., we assume that actions amongst agents are communicative acts. Agents can interpret a fellow agent's action when they observe them and learn which actions to take to convey a maximum of information to others [9]. As a result, agents learn when and what to communicate to each other to best achieve joint and individual goals.[10] From a technical perspective the major challenge was to extend the training procedure to explore in policy space and not – as usual – in action space, as one agent's belief about another agent's current state depends not only on the current state and observed action, but also on the policy explored.

First training results showed substantial instabilities in performance. Although training performance reached good performance levels, performance deteriorated massively when making slight changes to the agent team. Our analysis showed that agents learned "idiosyncratic"[11] behavior. After changing the training protocols to implement cross-play and league play schemes, results could be stabilized. As our preliminary results, discussed in more detail below, make clear, this also vindicated the resilience hypothesis presented above. Further investigation and training of the system is required to potentially learn optimal communication patterns and timing under low communication bandwidth constraints or electronic warfare (EW) conditions.

C. GhostPlay's Approach to Simulation

The in-process combat simulator is a central piece of the GhostPlay environment. The simulator orchestrates interaction among objects in a staged war gaming scenario. The simulator's computing power is key as scenarios with a fairly large number of entities need to be played quickly over several thousand time steps. Therefore, the simulator was built to be deployable "in-process" and directly interact with the objects to be trained without network latency. The simulator also has precautions to play scenarios in multiple time resolutions.

The simulator is extensible horizontally by adding new objects to the scenario and vertically by extending individual models with more details. While playing low resolution scenarios, the simulator works primarily with probabilistic models (representing summary statistics of interaction effects) and targets temporally extended scenarios as they would occur in an Anti-Access/Area Denial (A2/AD) situation. Equipped with higher resolution models, we have specified operational behavior down to level of modeling mechanical latencies of AAA turret movements or individual sensor control.

D. Preliminary Results

GhostPlay's preliminary results are encouraging. After less than one year of simulation-based research we observe that AD components behave in novel ways. New patterns reflect core tenants of the principles of war, as we argue below.

1) Single Platform Tactics

GhostPlay deliberately started out modelling a single air defense platform thereby using the FlakPz Gepard, a German AAA system, that is largely self-sustaining. The Gepard is also most qualified for the GhostPlay scenarios that require an AD system to operate on-board active sensors (search radar and tracking radar function) and passive sensors (optical periscope and infrared sensor) plus effectors while engaging targets on the move. We also wanted to experiment with different coordination policies to analyze, if platform behavior adapts commensurate with additional sensor and effector capabilities, as this would suggest that the platform was able to learn how to use additional technical capabilities. Therefore, we equipped the AAA with hypothetical additional sensor (e.g., infrared sensor) and effector capabilities in the simulation. We also wanted to know if the policies learned would take advantage of a fused Recognized Air Picture (RAP) using information from multiple sources and thus provided the system with a link-based, centrally supplied RAP.[12]

[8] Economy of effort suggests that it might not be economic to attack an artillery missile which costs US$150k with an AD missile that costs US$8m – unless the artillery missile may destroy an entity, which is an extremely important part of the defender's infrastructure. Disobeying economy of effort may quickly turn into massive losses of defensive capabilities and resilience.

[9] This is a rather radical standpoint. We expect that a real-world deployment will contain certain data aggregation and command nodes, however that individual systems will be able to work without them, but if they are available, make optimal use of them.

[10] The policy explored, which defines the behavior of agents, is assumed to be common knowledge to all agents in the team.

[11] For example, a AAA agent was observed to switch on its search radar. As the platforms have learned to operate mostly with passive sensors and networked RAPs, other agents believed this action to suggest the platform wants to signal that it is being attacked. This interpretation is not totally unintuitive. In general, however, switching on the search radar implies only that the platform wants to acquire more information about its close surrounding and does not automatically imply that the platform is under attack. If other agents do misinterpret such a behavior, they might move towards the sending platform to help, thereby giving up their position for no real reason.

[12] The RAP would be produced by a larger range data fusion process, using more powerful and longer-range sensors. As the RAP production involves processing and human validation of classification and identification, it provides a wider area view but may supper from reporting latency, to be considered when associating such information to local sensors.

This single-platform setup has produced a series of interesting findings that can be summarized as follows:

a) *The system learns sensor-control strategies to improve target classification.* Traditionally, for example, an AI classifier receives a video frame produced by the periscope camera to classify a target. In contrast, a periscope using the GhostPlay policy first learns how to zoom-in on a certain coordinate of interest as this leads to faster convergence on a stable classification.[13][14]

b) *The system learns multi-sensor control strategies:* The system is able to learn policies which implement situation-specific trade-offs between relying largely on passive sensors and deploying active sensors to minimize the risk of being detected and attacked by radiation-following missiles.[15]

c) *System learns to change behavior when a RAP is available:* In the same scenario the system behaves differently if it acts upon RAP-ensured situational awareness. Behavioral changes are most notable for the use of passive sensors (periscope). These sensors are mainly used for 360° searches if the global air picture is not available. For example, search directions focus on incidents when the platform needs the most time to turn due to mechanical constraints or to adjust the turret position early on to anticipate incoming threats.

d) *System learns to prioritize:* The system learns how to prioritize target engagements. We have used a swarm of 105 UAVs. The swarm flew in a pack formation and broke up shortly before the AD system to stage individual attacks. In the most demanding scenario 105 individual trajectories were meant to confuse the AAA sensors. At the moment the swarm broke up, a high-velocity threat approached the AAA system from a different angle. The system has mastered the challenge of, first, detecting the high velocity threat; second, recognizing that this threat is far more serious than the UAV swarm; and third, turning turret and weapon to engage this threat while the UAV swarm continues to perform fancy maneuvers.

e) *The fire-control policy learns appropriate timing.* The system learns policies, which discriminate between platforms that deploy weapons (e.g., attack helicopters or UCAVs) and loitering ammunition. Generally, the learned policy shows a tendency to engage loitering ammunition later and weapon carrying vehicles earlier if they are in reach of effectors.[16]

f) *The fire-control policy compensates low sensor resolution or track drops with UAV swarms:* Especially in attacks by smaller scale UAVs (e.g., attrition attacks) sensor systems and trackers may not be able to resolve each UAV individually or produce switching tracks and/or lose/drop tracks required to re-initialize. We observe that the learned fire control policy is comparably robust to these issues. The policy learns to engage a "pulk" with a series of coordinated barrage fire patterns to gradually reduce the swarm size, even when tracks have a comparably wide covariance. Should further tests vindicate this behavior, sensor quality would matter less to AAA systems, while opportunities to operate these systems with remote sensors (e.g., using sensors from other platforms or forward deployed sensors) would significantly increase.

Overall, we trained the AAA platform against a variety of different threats, ranging from single UAV/UCAV like the Bayraktar TB2, drone swarms, and helicopter attacks represented by Ka-52 and Mil Mi-28 combined with fast approaching missiles. Attackers were modeled with "local rule-based intelligence,", i.e., the attack pattern and objective were predefined with pre-specified approach trajectories. Attackers, however, operated on modeled rulesets prescribing how to respond to the detection of and the engagement by an AD system.[17] All models had associated a probabilistic damage model, which allowed realistic effector impact estimation on a target object, given the target's physical structure, effector type and impact area.

Preliminary results suggest that the AAA platform learns very fine-granular engagement tactics for different types of threats and even senses when it is important to destroy the target or only disable it. Compared to a traditional OODA workflow implementation, our system reduces the volume of ammunition required to protect assigned objects by around 12% vis-à-vis helicopters and up to 42% in scenarios against attacking swarms.[18] The project will extend and verify these figures further to publish detailed reports in the project's mid-term report. We plan to open the simulation environment and/or to establish a test bed, where vendors can compare their individual control strategies.

2) *Multi-Platform Tactics*

In addition to single-platform scenarios, we combined multiple AAA units of the same type to protect an airfield as a a scenario-specific high value asset (HVA). The intention of these training runs was to get first insights in what could be expected from having multiple AAA systems learning to team

[13] Similar to DeepMinds AlphaStar the action space is implemented with action macros, i.e., the system first determines the type of action (e.g. sensor control, effector-control, sensor number, all subsequent fields are then interpreted in context of the action macro).

[14] An interesting new opportunity is to connect the FlakPz with a passive sensor network (like TwinVis), with the passive system acting as a preliminary guidance and pre-warning system. Preliminary evidence suggests that this combination could greatly strengthen the survivability of the FlakPz as it reduces its electromagnetic emission.

[15] Evidence from some scenarios suggests that policies have learned to use the tracking radar to provoke the target to change direction in the attempt to escape the tracking beam. However, this needs to be analyzed in more detail, especially to ensure that observed behavior is stable and not just an unwanted artefact. This analysis will be done in the second project phase, where we plan to use a more elaborate aerial vehicle behavior in contrast to the current rules-based approach.

[16] As of now this is just an observation. We have not yet properly analyzed this behavior. But looking at scenario runs with platforms that carry weapons and use earlier generations of the trained policy suggests, that the earlier engagement may preempt the release of weapons by the platform. "As-early-as-possible" engagements also occur in scenarios with platforms that use models of laser-guided weapons, which could be interpreted as further evidence underpinning the observation.

[17] Following the principles discussed in [7] the local behavior in response to imminent threats and the orchestration of attacks while being engaged by the AD systems were modelled by Fuzzy inference but adapted for SEAD missions.

[18] In several UAV swarm scenarios, excess ammunition required by the OODA workflow-controlled systems was not the main issue. In these cases, the AAA platform simply did not survive the scenario.

up freely and without a central C2 coordination. We have achieved the following preliminary results:

a) *A group of AAA platforms learned to cooperate in defending against a drone swarm with 30 UAVs:* The AAA platforms' cooperative tactics was already rather complex (FIGURE 2): AAA platform 1 used its active sensors, while platforms 2 and 3 were observing the situation under emission control (EmCon). As 10 UAVs separated from the swarm to engage AAA platform 1, 20 UAVs proceeded further to the airfield as the main target. While AAA 1 engaged the attacking UAV swarm, AAA 2 attempted to sneak in by the main swarm. Meanwhile AAA 3 pretended being a "lame duck." Shortly before the UAV swarm staged its attack AAA 2 and 3 simultaneously engaged the main swarm. It turned out that the move of AAA 2 created a situation that severely restricted the freedom of maneuver of the swarm, which could be effectively neutralized. In 30% of all scenarios in which the AD systems had not been using this policy, the swarm prevailed and damaged the airfield significantly. By contrast, the AAA team using this policy outperformed the swarm in 98% of all scenarios played and protected the airfield.

FIGURE 2: AAA PLATFORMS (BLUE) DEFEND AIRFIELD AGAINST AGGRESSOR SWARM (RED). VIDEO MATERIAL GHOSTPLAY.

b) *Increasing survivability by re-organization:* In further tests we STARTED to investigate the effects of attrition attacks. We used 10 AAA systems to protect a HVA against an aggressor swarm of 105 UAV. The 10 AAA systems were positioned around the HVA. The attacking swarm leveraged the principle of "relative strength," which means that 70-90 UAVs would concentrate on a geographically small area, creating an overwhelming force for the two or three AAA systems deployed in that region. In parallel smaller UAV swarms would try to distract AAA systems and keep them busy in their positions. The concentrated force led to unavoidable losses of AAA systems in the scenario. In earlier training stages the AAA solution was lost and the number of UAVs that survived was large enough to attack the HVA. At later stages of the policy, the AAA platforms learned to continuously re-organize group assignments and re-prioritize targets. Consequently, whenever one AAA system was becoming dysfunctional another AAA platform was moving in (even preemptively, when a AAA platform was running low on ammunition), such that the UAV swarm was substantially decimated and no longer able to substantially harm to the HVA. Overall, the policy suggests that in 9,864 out of 10,000 scenarios a constellation of 10 AAA was able to put up effective protection against a 105-member UAV swarm, losing not more that 3 AAA platforms.

We need to emphasize that these are early preliminary results based on idealized assumptions. For example, the scenarios assume that AAA platforms share internal status information among them and can freely choose sensor deployment without having to adhere to EmCon rules. In addition, the AAA platforms did not have any restrictions to move out of their positions, nor were they bound by rules of engagement (RoE). We will use further simulation runs to explore how different RoE will affect the freedom of maneuver of the AAA platforms under consideration. We will also scrutinize how RoE need to be crafted to ensure effective human control, without preventing the AAA platforms from delivering the results already accomplished.

In addition, there are several technical caveats. On the one hand it is by no means certain, that operating multiple AAA platforms in a federation without central coordination [19] would produce any meaningful results. On the other hand, all AAA platforms could just jump on the same target as soon as it is in reach of their effectors, thus using available capacities very inefficiently. Although our preliminary results are very encouraging, we have taken precautions to learn stable policies thereby using team rotations, "other-play" and learning protocols like league-updates. Further research is needed to ensure, that policies do not learn to agree on implicit communicative acts, which would break the POMDP conditions and may lead to instable behavior. Given imperfect perception models and simulated "fog-of-war" effects, the latter may be substantially more difficult as compared to computer games and will require further efforts in upcoming project phases.

3) Summary

Initial findings suggest that learned policies can create behavioral patterns that reflect key principles of war. More fine-granular control of the sensor-effector network reduces the amount of force required (economy of force) to establish effective protection (objective). Our AAA systems advance situational awareness at the platform level by considering more information than only kinematic aspects of the target object. This enables the platform to anticipate adversarial moves and enables emergent and adaptive countermeasures. This behavior will make it impossible for the adversary to "read" and understand the AAA system. Thus, air defenders can exploit newfound elements of surprise that shift the initiative to their benefit.

Defense systems leveraging network-centric warfare mainly focus on building federated solutions by seamlessly integrating all components. Our preliminary results show that cognition is about to significantly augment these federations as every component can interpret the current and future behavior of its companions based on policies known by all elements of the federation. This would offer new ways to ensure resilience in non-benign environments in which communication is likely to be missing and data will be corrupted.

[19] In practical deployments there is at least an assigned sector of responsibility in which the platform actively engages targets.

III. Balancing Ethics with Performance

In democracies armed forces operate within a framework set by ethical and moral principles as well as the rule of law.[20] Within this framework, armed forces will sooner, rather than later, grapple with technical autonomy. In this context, defense AI causes significant concerns as it serves the use of force. Therefore, if defense AI is used to defend democracies, it must necessarily be embedded in national legal and value-oriented frameworks, relevant supranational rules and international law. In practice, however, it proves to be a major challenge to incorporate legal, ethical, or societal norms into the functions of AI systems.

Since 2017, governmental and non-governmental organizations have produced lists that outline generic, mandatory quality attributes for AI systems. These lists can be seen as a first attempt to combine ethical, legal, societal, and technological considerations. But these lists are far from sufficient to realize or promote core values such as human dignity and freedom, peace and justice, or soldierly virtues such as love of one's homeland, truthfulness, or courage.

Thus, a key research aspect of GhostPlay is to evaluate the applicability of the new IEEE 7000™ 2021 standard for Value-Based Engineering (VBE), which became effective in September 2021. Ideally, applying this standard would lead to defense solutions, with different qualities. That's why GhostPlay wants to consider the entire universe of values that German Armed Forces attribute to their leadership principle of Innere Führung. Being in close contact and contributing to the standard's further evolution, GhostPlay will be the first defense AI application worldwide designed to fully comply with IEEE 7000™-2021.

Moreover, GhostPlay's use of the IEEE 7000TM standard will produce learning materials to train Value Leads, a new job description in systems development. These Value Leads possess the philosophical and technical understanding required for VBE with the goal to make Germany a pioneer and leading nation educating value-sensitive defense AI engineers and developers.

IV. GhostPlay's Approach to Innovation Management

GhostPlay is a capability and technology development project that uses cutting-edge insights from academic as well as applied research to provide the Bundeswehr with a novel level of tactical versatility. Tactical versatility complements the Bundeswehr's strive for information, decision, and effects superiority. GhostPlay's key added value stems from the fact that context and consequence-aware solutions can be transferred across applications used in different military domains. This creates valuable opportunities for cross-pollination between domains and military services.

Ultimately, GhostPlay's demanding development agenda requires an innovation management approach that is agile and holistic. To this purpose GhostPlay combines the Real-Option approach [8] with an agile development process. With this approach new research and implementation topics are assessed according to their expected operational value, adopting a hypothetical pricing scheme, like financial option pricing. As the scheme takes into account internal and external risk factors, it balances risks and opportunities to maximize the expected operational value, which can be created by the assigned budget.

Practically, the innovation portfolio is evaluated every six months, combining external information about recent conflicts, technology trends and recent initiatives of Western forces, collected and organized by the Defense AI Observatory, with actual market requirements as perceived by industry partners and the proper findings of our research project.

V. Outlook

After one year, GhostPlay has delivered encouraging results that underline the feasibility and potential improvements of tactical AI and emergent coordination in an AD environment.

Already at this stage, GhostPlay's project partner Hensoldt has decided to transfer the project's sensor resource management capabilities into a new environment to coordinate the deployment of active and passive sensors with tactical AI. This will create new capabilities for armed forces and vindicates the project's methodological and technological approach.

Moreover, we will extend the set of principles used for GhostPlay by replacing a fuzzy logic-based interference mode currently used for attacking systems. At the next stage attackers shall use the same tactical AI and emergence principles to develop new and change existing tactics during a mission in a "counter-play" training protocol. This means that AD systems and SEAD systems would be trained in alternating cycles. Whenever a more successful AD policy is found, SEAD polices will be adapted to overcome the new AD policy and vice-versa. These "opponent-play" cycles will enable both sides to continuously learn increasingly fine granular and complex behavior, enabling them to cope with today's most dangerous threats at a certain stage.

AI-based SEAD tactics are of specific interest, as they directly address a current capability gap. To develop SEAD tactics against sophisticated threats such as S400 and S500 AD compounds, the existing AD capability will be complemented with surface-to-air-missile models, which may extend the purely reactive RL architectures used today with planning in large scale POMDP methods.

While many aspects of GhostPlay still require further research and analysis in terms of robustness and effects before they could enter operational service, the project creates added value for different military tasks:

a) *Non-traditional red teaming for future force planning:* The GhostPlay simulator and AI models can be used to test new sensor/effector constellations. GhostPlay provides the first environment, in which AI methods learn how to best use available physical capabilities. This provides force planners with advanced insights on how new sensors, effectors, communication, and platform capabilities would affect future tactics. The system can thus be used to find the best capability combinations and efficiently develop operational requirements for new systems.

[20] As industry is actively researching aspects of technical autonomy, for example, to support autonomous driving and robot assistance, there is no plausible reason, why these technologies would not show up in a military context. Thus, the effects and possibilities of such capabilities must be understood and analyzed both, in terms of future force planning and in terms of potential future threats.

b) *Non-traditional red-teaming for projects currently under development:* GhostPlay can provide a testbed for system vendors to test their concepts and tactics against a hard-to-predict adversary. Currently new systems are evaluated against scenarios and vignettes developed by military analysts, but the selection of scenarios is biased towards allied doctrine and allied thinking on expected adversarial behavior. By contrast, GhostPlay operates "model-free" and learns tactical behavior without any preselected vignettes. This approach provides behavioral patterns not yet seen in practice or in existing models and thus augments existing testbeds, better prepare allied systems, and potentially uncovers unknown weaknesses in systems under development.

c) *Non-traditional red-teaming for crew-training:* Being setup in a DIS (IEEE) framework, GhostPlay components can be integrated in pilot and air defense simulators to train crews on yet unseen tactics.

d) *Transferring GhostPlay to other domains and mission areas:* GhostPlay's approach and policies can underpin the development of defense solutions meant to coordinate complex intercept missions without a central C2 component. This could provide novel solutions to protect naval platforms against surface and subsea threats and could enhance solutions to provide Joint Tactical Fire Support (JTFS), for example.

Finally, GhostPlay partners also mull the idea of potentially operating a "GhostPlay light" environment, i.e., a digital twin of the simulation environment with lower fidelity and unspecific sensor models. "GhostPlay light" could be hooked up with commercial video games. The intention is to leverage the "wisdom of the crowd" by involving many professional, semi-professional and hobby pilots to detect new and unconventional tactics. These new tactics could then be used to confront and refine GhostPlay. The respective results could be transferred into the restricted simulation environment of armed forces, which operate realistic sensor models.

ACKNOLODGEMENT

This Paper is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr which we gratefully acknowledge [GhostPlay].

LITERATURE

[1] Work, Robert O. and Shawn Brimley, 20YY. Preparing for War in the Robotic Age, Washington, DC, Center for a New American Security, 2014.

[2] Science and Technology Trends 2020-2040. Exploring the S&T Edge, Brussels, NATO Science and Technology Organization, 2019.

[3] Horowitz, Michael C. and Shira Pindyck, "What is a military innovation and why it matters," Journal of Strategic Studies, 2022, https://doi.org/10.1080/01402390.2022.2038572 [Last access 31 August 2022].

[4] Borchert, Heiko, Torben Schütz, Joseph Verbovszky, Beware the Hype. What Military Conflicts in Ukraine, Syria, Libya, and Nagorno-Karabakh (Don't) Tell Us About the Future of War, Hamburg, Defense AI Observatory, 2021, https://defenseai.eu/daio_beware_the_hype [Last access 31 August 2022].

[5] UK Defence Doctrine, Joint Doctrine Publication 0-01, Shrivenham, Development, Concepts and Doctrine Center, 2014.

[6] Ernest, Carroll, Lee, et.al, Genetic Fuzzy based Artificial Intelligence for Unmanned Combat Aerial Vehicle Control in Simulated Air Combat Missions. DOI:10.4172/2167-0374.1000144 [Last access 31 August 2022].

[7] İşci, Hasan, Gülay Öke Günel, Fuzzy logic based air-to-air combat algorithm for unmanned air vehicles, International Journal of Dynamics and Control, 10:1 (February 2022), https://link.springer.com/article/10.1007/s40435-021-00803-6 [Last access 31 August 2022].

[8] Trigeorgis, Lenos, Real Options. Managerial Flexibility and Strategy in Resource Allocation, Cambridge, MIT Press, Cambridge, 1999.

[9] Foerster, Jakob N. et. al., Learning to Communicate with Deep Multi-Agent Reinforcement Learning, archiveX, https://arxiv.org/abs/1605.06676 [Last access 31 August 2022].

[10] Uppal, Rajesh, "Israel successfully tests multilayered air defense system to defend against barrage of short-, medium-, and long-range ballistic missiles," IDST, 15 May 2022, https://idstch.com/geopolitics/srael-successfully-tests-multilayered-air-defense-system-to-defend-against-barrage-of-short-medium-and-long-range-ballistic-missiles/ [Last access: 31 August 2022]

[11] Sun, Yi, Daan Wiestra, Tom Schaul, and Jürgen Schmidhuber, "Stochastic search using the natural gradient," ICML 09: Proceedings of the 26th Annual International Conference on Machine Learning, 2009, https://dl.acm.org/doi/10.1145/1553374.1553522 [Last access: 31 August 2022]

[12] Wiestra, Daan, Tom Schaul, Tobias Glasmachers, Yi Sun, Jan Peters, Jürgen Schmidhuber, "Natural Evolution Strategies," Journal of Machine Learning, 2014, https://jmlr.org/papers/v15/wierstra14a.html [Last access: 31 August 2022]

[13] Graves Alex. et.al., "Hybrid computing using a neural network with dynamic external memory,", Nature, 20 October 2016, https://www.nature.com/articles/nature20101 [Last access: 31 August 2022].

[14] Kanerva, Pentti, Sparse Distributed Memory, Boston, MIT Press, 1988.