Trends in Machine Learning for Adaptive Automated Forces

LTC Austin Starken

US Army

Orlando, FL

austin.starken@us.army.mil

Dr. Sean Mondesire, Dr. Annie S. Wu, Dr. Bruce Caulkins

University of Central Florida

Orlando, FL

sean.mondesire@ucf.edu, aswu@cs.ucf.edu, bruce.caulkins@ucf.edu

ABSTRACT

Critical to Army readiness, simulation-based training offers a cost and time-effective way to keep personnel well-versed in their roles, responsibilities, tactics, and operations. Simulation-supported exercises currently require long planning timelines and significant resources. Although semi-automated military simulations provide essential behavioral artificial intelligence to assist in fulfilling participant roles, they still need human simulation operators to control friendly and opposing forces. Exercise support simulation operators come directly from the intended training audience, assigning Soldiers role-playing duties versus training with their organization. Units train with a fraction of their team, reducing training quality and its overall impact. One method for reducing overhead and improving the quality of simulation-supported training is implementing fully automated and adaptive opposition forces (OPFOR).

DeepMind's AlphaStar, AlphaZero, and MuZero illustrate the progression of machine learning research. Using large datasets or generalized algorithms, these agents learned how to play and defeat professional players at complex, combative strategy games. These games include delayed and sparse rewards, imperfect information, and massive state spaces, all feats that support the idea that machine learning may be the key to developing adaptive OPFOR in constructive military simulations.

This paper surveys the existing literature on the use of machine learning for automated OPFOR decision-making, plan classification, and agent coordination. This analysis serves as a starting point for future research on the current capabilities and limitations of developing adaptive OPFOR in support of constructive military simulations.

ABOUT THE AUTHORS

LTC Austin Starken is a Functional Area 57 (Simulation Operations) Army officer with 17 years of experience training Soldiers. He served as a mission training complex (MTC) Director, technical integrator of the Army Battle Lab Collaboration Simulation Environment, and the operations officer for the Joint Cyber Training Enterprise.

Sean Mondesire, Ph.D. is an Assistant Research Professor of Big Data Analytics and Cybersecurity at the University of Central Florida's School of Modeling, Simulation and Training (SMST). Dr. Mondesire's research interests are machine learning and big data analytics of large-scale, dynamic training systems.

Annie S. Wu, Ph.D. is an Associate Professor of Computer Science and Director of the Evolutionary Computation Lab at the University of Central Florida. Before joining UCF, she was a National Research Council Postdoctoral Research Associate at the Naval Research Laboratory.

Bruce D. Caulkins, Ph.D. is an Associate Research Professor at the University of Central Florida's School of Modeling, Simulation, and Training (SMST). He is also the Director of the M&S of Behavioral Cybersecurity Graduate Program at UCF. He is a retired Army colonel with over 28 years of active-duty service, culminating as the Chief of the Cyber Strategy, Policy, Plans and Exercises Division, J65, US Pacific Command.

Trends in Machine Learning for Adaptive Automated Forces

LTC Austin Starken
US Army
Orlando, FL
austin.starken@us.army.mil

Dr. Sean Mondesire, Dr. Annie S. Wu, Dr. Bruce Caulkins
University of Central Florida
Orlando, FL
sean.mondesire@ucf.edu, aswu@cs.ucf.edu, bruce.caulkins@ucf.edu

INTRODUCTION

National threats continue to challenge local and global safety and security (Department of Defense [DoD], 2022). The complexity of future warfare grows as threatening nations and organizations straddle the tenuous threshold of competition and conflict on a multi-domain battlefield (Army Futures Command [AFC] Futures Concepts Center [FCC], 2020). To deter the aggression of perceived threats, United States (US) military forces must rapidly develop and maintain combat readiness (DoD, 2022).

The Army provides the DoD with trained and ready forces through training that is "challenging, relevant, realistic, and performed to the highest standards" (Headquarters, Department of the Army [HQDA], 2021). Ideally, training takes place in live environments at the individual to company level, supported by *training aids, devices, simulations, and simulators* (*TADSS*). However, higher echelon units at the battalion, brigade, and above have training requirements that live environments may not support due to time, cost, or safety. Higher echelon units require an *integrated training environment* (*ITE*) of *live, virtual, and constructive* (*LVC*) training (HQDA, 2021, Appendix J). Live training consists of real Soldiers operating real systems; virtual training includes real Soldiers operating simulated systems; while constructive simulations require real Soldiers to operate numerous simulated entities. The Army defines entities as "independent [simulated] objects with complex behaviors and attributes (e.g., personnel, vehicles, complex munitions, and key communications devices)" (United States Army Combined Arms Center [USACAC], 2018, p. 5). Through an appropriate mix of LVC training, Army units train and maintain the proficiency of their formations within a "band of excellence" (HQDA, 2021, pp. 1-3).

TADSS and an ITE enhance training and reduce costs but still require considerable overhead. The Army needs technical simulation, network, and infrastructure *subject matter experts* (*SME*) to operate and integrate training simulations. Simulated wraparound forces (higher and lower echelon blue forces, neutral organizations and civilians, enemy forces, etc.) require personnel to develop operational and tactical plans and provide control during execution. A training exercise's operational, technical, logistical, and administrative requirements result in 12-24 months of planning (Joint Chiefs of Staff, 2015, pp. E-3). The added complexity of developing relevant scenarios that account for *Multi-Domain Operations (MDO)* (AFC FCC, 2020) and *dense urban environments* will only lengthen planning timelines and complicate efforts. To summarize:

> The Integrated Training Environment (ITE) Training Aids, Devices, Simulators (TADSS) currently lack the ability to allow units and Soldiers to conduct realistic, multi-echelon, collective training, seamlessly from squad to Army Service Component Command (ASCC) echelons, anywhere in the world, and require significant training overhead (time, money, people) to utilize. (USACAC, 2018, p. 1)

Currently, the Army is developing its future training simulation system, the *Synthetic Training Environment* (*STE*) (USACAC, 2018). The STE is the Army's attempt to centralize and standardize Army simulations to simplify the exercise development process allowing commanders to spend more time training and less time planning. It aims to reduce the technical and logistical overhead commonly associated with simulation-supported (sim-supported) exercises using advancements in simulation technology. To achieve the STE's Critical Operational Attributes (USACAC, 2018, p. 16), the Army acknowledges the need to leverage *artificial intelligence* (*AI*) and *machine learning*

(*ML*) technologies (Kimmons, 2020; Rozman, 2020; USACAC, 2018; 2019). Entities or agents within a simulated environment capable of adapting to the training audience can enhance learning, performance, and engagement (Van Den Bosch et al., 2020).

This paper identifies requirements to develop intelligent agents that can make decisions, recognize opponent plans, and coordinate actions in response to the training audience. This paper analyzes recent advancements and trends in building adaptive automated forces, identifies current challenges the training community has with realizing automated *opposition forces* (*OPFOR*), and discusses potential innovative paths to overcome those obstacles for Army training simulations. This work provides a background of the current state of Army constructive and virtual simulations and establishes a baseline understanding of ML concepts for adaptive agents. Next, we discuss recent advancements in ML that support agent decision-making, plan recognition, and multi-agent coordination. Finally, the paper discusses what these advancements mean for the Army, what research gaps remain, potential solutions to these challenges, and concludes our findings.

BACKGROUND

Current Military Simulations

Army training simulations primarily include virtual and constructive simulations. Virtual simulations consist of a complete physical mockup of a weapon system that engages targets in a virtual world, such as the Army's *Aviation Combined Arms Tactical Trainer* (*AVCATT*) (United States Army Acquisition Support Center, 2022). Or they can be as simple as a 3D first-person shooter, such as the Army's *Engagement Skills Trainer* (Program Executive Office for Simulation, Training and Instrumentation [PEO STRI]). Virtual trainers also include virtual reality and augmented reality technologies, as in the STE's planned *Integrated Visual Augmentation System* (*IVAS*) *Squad Immersive Virtual Trainer* (*SiVT*) (PEO STRI). Constructive simulations best support higher echelon exercises by simulating large formations and operational and strategic assets not typically available in a live home-station training event. *Joint Semi-Automated Forces* (*JSAF*) is a simulation used for joint-level training and experimentation for tactical to strategic operations. *One Semi-Automated Forces* (*OneSAF*) is the Army's central constructive simulation that supports Army training, test and evaluation, analysis, intelligence, acquisition, and experimentation communities (PEO STRI). OneSAF and JSAF model individual entities up to brigade formations (Padilla, 2012, pp. 853-857). Table 1 provides a small subsection of the virtual and constructive simulations used across the joint force.

Table 1. US Military Simulations, adapted from (Padilla, 2012, p. 867)

Name	Type	User	Purpose	Operational Level	Military Unit
AFSERS	Constructive, Virtual	USAF	Training & Mission Rehearsal	Tactical	Battalion
AWSIM	Constructive	USAF	Training, Mission Rehearsal, Experimentation	Operational, Tactical	Wing
JCATS	Constructive	US Joint	Training, Analysis, Experimentation	Tactical	Up to Battalion
JSAF	Constructive	US Joint	Training & Experimentation	Strategic, Operational, Tactical	Up to Brigade
MTWS	Constructive	USMC	Training & Analysis	Tactical	Up to MEF and JTF
OneSAF	Constructive	Army	Training, Experimentation, & Acquisition	Strategic, Operational, Tactical	Up to Brigade
VBS	Constructive, Virtual	USA	Training & Mission Rehearsal	Tactical	Squad
WARSIM	Constructive	USA	Training & Mission Rehearsal	Operational, Tactical	Brigade and above
RESA	Constructive	USN	Training & Acquisition	Operational, Tactical	N/A

Organizations: USA: US Army, USAF: US Air Force, USN: US Navy, USMC: US Marine Corps
Simulations: AFSERS: Air Force Synthetic Environment for Reconnaissance and Surveillance Mode, AWSIM: Air Warfare Simulation, JCATS: Joint Conflict and Tactical Simulations, JSAF: Joint Semi-Automated Forces, MTWS: Marine Ground Task Force (MAGTF) Tactical Warfare Simulation, ONESAF: One Semi-Automated Forces, VBS: Virtual Battlespace, WARSIM: Warfighters Simulation, RESA: Research, Evaluation, and System Analysis
Units: MEF: Marine Expeditionary Force, JTF: Joint Task Force

With recent advancements in AI and ML, researchers are witnessing the potential of adaptive computer-generated agents (Ballanco, 2019; Fossaceca & Young, 2018; Priya Narayanan, 2021). However, current AI and ML methods in DoD simulations are still semi-automated and rule-based (Abdellaoui et al., 2009; Oswalt et al., 2019), thus requiring human simulation operators (sim-operators) to control *computer-generated forces* (*CGF*). JSAF/ONESAF entities can fully automate their actions but need a mission plan. A sim-operator must designate the task and purpose (mission) of every entity or unit and any path(s) they should follow before the simulation starts. Once running, the simulated entities can sense and react to their surroundings based on a limited number of behaviors. Even with these features, JSAF/OneSAF entities cannot handle novel situations outside their pre-defined behaviors. Sim-operators must monitor and intervene when necessary, thus requiring training in basic military tactics and how to operate the simulation to ensure that the presented agent behaviors make sense to the training audience (realism). As an exercise grows, so do support requirements and administrative overhead.

The government and industry are aware of current military simulation AI limitations and are attempting to improve them. Bohemia Interactive recently released their new *VBS® Control Behavior Pack 1*, allowing users to build customizable behaviors (Bohemia Interactive Simulations, 2020). The behaviors are pre-programmed based on what Bohemia identifies as typical use cases. Also, exercise planners or SMEs must pre-construct the behavior trees before execution (Bohemia Interactive Simulations, 2020). Once the training event executes, the agents cannot operate outside their pre-defined behavior trees, preventing proper adaptation to the trainee's actions.

The *US Army Training and Doctrine Command Proponent Office for Constructive Simulations* works with Army OneSAF user communities (training, test and evaluation, analysis, intelligence, acquisition, and experimentation) to constantly improve OneSAF and maintain its concurrency through a requirements prioritization process (USACAC, 2005). In addition, organizations can make slight modifications to local copies of their software depending on their environment. Although this allows for a constantly improving simulation, the behaviors are simplistic and not adaptable.

AI in Army simulations has matured over the years but is rigid, predictable, and only adaptable through human intervention. It may be possible to reduce administrative burdens such as sim-operator training and management by creating CGFs that can adapt to trainee tactics.

Artificial Intelligence and Machine Learning for CGFs

Game developers have experimented with AI for decades to develop intelligent agents. Techniques include finite state machines, behavior trees, fuzzy logic, Markov systems, goal-oriented behaviors, and various combinations of two or more methods (Millington & Funge, 2009). Monolith's commercial horror game *F.E.A.R.* implemented *goal-oriented action planning* (*GOAP*) to produce the appearance of realistic tactical behaviors in real-time between agents who were unaware of one another (Orkin, 2006). Guerrilla Games' Horizon Zero Dawn utilized hierarchical task networks to control individual local agents, local and global agent herds, and a blackboard to coordinate individual agent actions (Thompson, 2019). Even though agent intelligence and coordination are improving in today's commercial video games, they still cannot learn a player's skill level or unique tactics and adapt.

One potential method for preventing trainees from exploiting AI limitations is to enable agent adaptation through ML. Four of the most prolific methods of ML include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. *Supervised learning* leverages large, labeled datasets to train agents. Researchers used this process to clone human player behavior in the 3D first-person shooter *Counter-Strike Go* through pixel data scraped from human play sessions (Pearce & Zhu, 2021). Supervised learning also enabled AlphaStar to defeat 80% of ranked players in *StarCraft II* (*SC2*) (Vinyals et al., 2019). *Unsupervised learning* trains machines by providing them with large amounts of unlabeled data and allowing them to discover patterns. Unsupervised learning may not fit well with military simulation environments due to potential noise in the data and the production of undesired behaviors (Fossaceca & Young, 2018; Roessingh et al., 2017). However, unsupervised learning methods have anticipated player tactics in a video game environment resulting in improved human-agent cooperation (Chen, 2017). *Semi-supervised learning* is a hybrid of both approaches. In this method, a machine trained on a set of labeled data develops a predictive model. Then, the trained machine receives unlabeled data producing pseudo-trained data. By keeping only accurately labeled pseudo-data and adding it to the machine's baseline, the machine iteratively improves its ability to classify future unlabeled data. Figure 1 illustrates this process. Facebook recently employed semi-supervised learning to improve its speech recognition algorithms (Kahn et al., 2020). Although unsupervised learning may not handle the noise of military simulations, the iterative nature of semi-supervised learning could help filter noisy data and build confidence in the application technique through observed iterative improvements during training. *Reinforcement* *learning* (*RL*) requires no initial data and enables an agent to learn and operate in its environment through a trial-reward process (Lee et al., 2018).

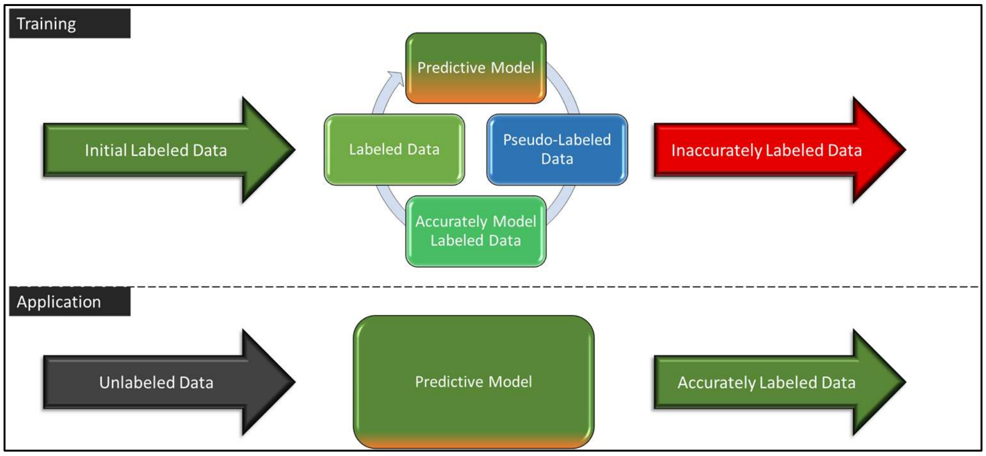

Figure 1. Semi-Supervised Learning

Different combinations of supervised, unsupervised, and reinforcement learning approaches have achieved remarkable results in developing agents capable of operating in large state spaces. Computers are capable of learning and mastering games in a myriad of ways. They can map raw pixel data to control inputs in simple video games (Mnih et al., 2016; Mnih et al., 2015). Use self-play to determine the optimal move in any situation in perfect information board games (Silver et al., 2018). Or even train against other agents designed to directly exploit their weaknesses (Vinyals et al., 2019). Adaptive CGFs in military simulations must be capable of making decisions, recognizing the trainee's plan or intent, and coordinating actions. In the following sections, we discuss the state of these capabilities and how they contribute to the development of adaptive OPFOR.

MACHINE LEARNING

Military training simulation environments consist of significantly high dimensions and variables, making it near impossible for tactical and technical SMEs to develop rule-based solutions that consider every possible situation. Advances in ML and deep learning algorithms are making headway in handling large complex environments similar to military simulations. The Army can potentially use these algorithms in Army simulations to produce intelligent OPFOR capable of adapting to each organization's skills and techniques. To improve Army simulations, agents must adapt their decision-making, plan and intent recognition, and coordination to truly challenge Army Soldiers and teams.

Automated Agent Decision-Making

Policies govern intelligent agent decision-making by defining what action an agent should take based on its current situation. Similarly, Army doctrine dictates that squads receiving fire from an enemy bunker should execute *Battle Drill 5: Knock Out a Bunker* (HQDA, 2016, pp. J-15). As it is impossible to define every possible action an agent should take in every likely scenario in a training simulation, the Army cannot define every action a Soldier should take in battle. Thus, the Army trains a core set of tactical tasks during initial training, and Soldiers expand their knowledge by training with their units. Soldiers grow their foundational knowledge through exposure to additional training scenarios, and this experience improves Soldier adaptability. Just as Soldiers learn to make critical decisions in different scenarios through training, intelligent agents can learn to make decisions through RL.

RL agents learn how to make appropriate decisions through trial and error in an environment that grades their performance with a reward. RL methods are either value-based or policy-based and may or may not depend on a local environmental model (Lapan, 2020). Value-based methods use a value function to define the anticipated value of state-action pairs. At the same time, policies explicitly tell an agent what action to take given a specific state. *Q-Learning* (Watkins & Dayan, 1992) is an early value-based RL algorithm that is the foundation of many advanced techniques today. Policy-based methods typically use policy gradients to discover the optimal parameters to generate a policy that maximizes the agent's cumulative reward (Lapan, 2020, p. 286). In contrast, model-based methods build a local model to predict or plan the next best action to take. These methods can work in isolation or leverage the strengths of each approach. However, the true advantage comes when combining RL methods with *neural networks* (*NN*).

Two recent policy optimization algorithms are *Trust Region Policy Optimization* (*TRPO*) (Schulman et al., 2015) and its updated version *Proximal Policy Optimization* (*PPO*) (Schulman et al., 2017). TRPO and PPO improve the stability of policy optimization by preventing an agent from diverging too far from its current policy. The algorithm only updates the agent's action if the change is within a defined boundary. PPO is much simpler to implement and has shown positive results in robotic locomotion and Atari games (Schulman et al., 2017). PPO served as a target selection method in a larger framework designed to control agents in large state spaces (Shen et al., 2021) and has been used in multi-agent systems to facilitate cooperation amongst agents (Yu et al., 2021).

Actor-critic algorithms leverage the best value-based and policy optimization methods by employing a critic that uses a value function to govern the parameters of an actor's policy (Grondman et al., 2012). The *asynchronous advantage actor-critic* (*A3C*) employs multiple actor-critic pairs that work in parallel and has shown rapid learning gains while minimizing computing resource requirements (Mnih et al., 2016). When coupled with the ability to maintain direct memory of its experience, such as with *long short-term memory* (*LSTM*) (Hochreiter & Schmidhuber, 1997), A3C has been used to develop intelligent military CGFs (Toghiani-Rizi et al., 2017) and has supported the development of AlphaStar (Vinyals et al., 2019).

Deep reinforcement learning (*DRL*), Figure 2, pairs RL algorithms with NNs to develop a policy through environmental rewards and state observations (François-Lavet et al., 2018). Deep Q-learning implements a *deep Q-network* (*DQN*) in place of a *Q-table* to approximate all the *Q-values* for every action in a given state (Mnih et al., 2015). DQN has seen iterative improvements with the addition of short-term memory (Kapturowski et al., 2019), episodic memory, incentivized exploration (Badia, Sprechmann, et al., 2020), and the balancing of exploration versus delayed rewards (Badia, Piot, et al., 2020). These enhancements resulted in agents handling game environments that provide multiple immediate rewards to significantly delayed and sparse rewards, as found in 57 different Atari games (Badia, Piot, et al., 2020).

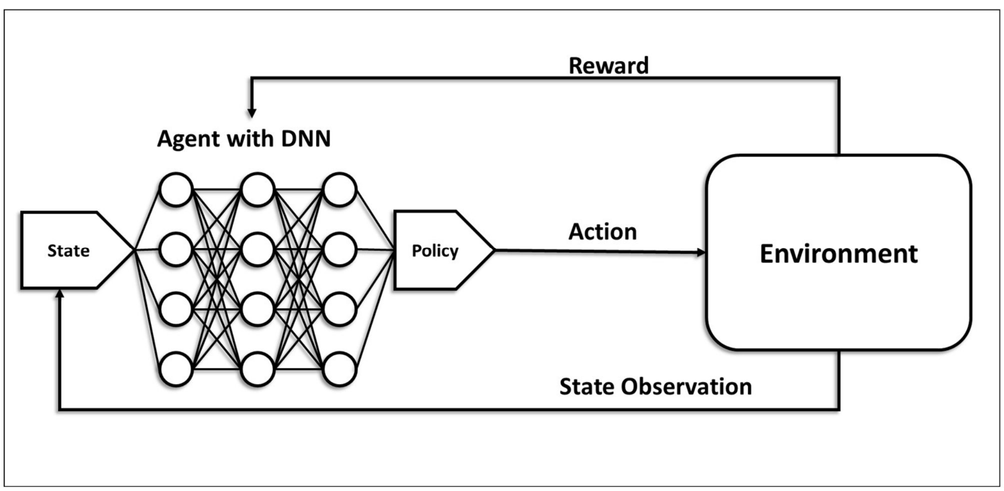

Figure 2. Deep Reinforcement Learning

In addition to mastering Atari games, DRL has led to agents capable of operating in complex environments such as real-time strategy games. Using a *convolutional neural network* (*CNN*) to process visual data, AlphaStar used supervised learning of player data, DRL, and *multi-agent reinforcement learning* (*MARL*) to learn how to master SC2 (Vinyals et al., 2019). By pitting AlphaStar against different versions of itself, including those specifically designed to exploit AlphaStar's weaknesses, as well as learning from matches against human opponents, AlphaStar achieved grandmaster status in the global rankings for SC2 (Vinyals et al., 2019). Though impressive, AlphaStar requires significant human player data to learn SC2 and cannot transfer its skills outside the game. AlphaZero, on the other hand, is a more generalized agent that taught itself how to play and master chess, shogi, and Go (Silver et al., 2018). AlphaZero also used a CNN to process visual data but used a generalized RL algorithm and a generalized Monte Carlo Tree search to learn and master each board game by repeatedly playing itself (Silver et al., 2018). MuZero takes AlphaZero even further by having its NN dynamically create its own environment model to plan its next move, enabling it to play both board games and Atari games using the same algorithm (Schrittwieser et al., 2020). These advancements illustrate the power of DRL to create intelligent agents capable of adapting to an environment with or without large data repositories. Though the Army generates significant training data through sim-supported training, the policies and infrastructure required to leverage this data can be complex. However, general intelligence in virtual environments is quickly evolving and could lead to adaptive computer-generated OPFOR without needing large, labeled datasets.

Plan Recognition and Agent Coordination

The military invests significant resources in intelligence systems to determine the plan and intent of enemy forces. Even with these investments, predicting the opposition's course of action at any given moment is challenging. Military simulations have access to all real-time information at any given moment. However, leveraging this advantage produces unrealistic actions that reduce realism and cause negative training. Further, access to simulation data does not provide insight into what a training audience has planned or what their ultimate objective might entail. Adaptive OPFOR must be capable of recognizing the plans or goals of other agents in military simulations. Predicting agent intent is crucial for enabling coordination amongst OPFOR agents and adapting their tactics in response to the actions of the training audience.

Plan, activity, and intent recognition (PAIR) research allows intelligent agents to recognize the actions of other agents or humans, determine why they are conducting those actions, and determine the observed agent's next move. A *plan* is a sequence of steps that enable an agent to accomplish a goal (conduct movement to contact), *activity* is a series of actions a person or agent is currently conducting (squad moving in column formation through enemy-controlled territory), while *intent* is the goal of the agent(s) (locate and engage enemy forces). Agents capable of observing the action of adversarial agents and recognizing their intent is critical to the development of adaptable OPFOR.

Current approaches to PAIR employ traditional ML and brain-inspired strategies but mostly center around logic-based reasoning and deep learning techniques (Van-Horenbeke & Peer, 2021). Logic-based methods use one of two techniques. A pre-defined plan library allows an agent to compare observations against known plans (Avrahami-Zilberbrand & Kaminka, 2014; Ni et al., 2021). However, *plan recognition as planning* instills an agent with domain knowledge which it uses to develop hypotheses to determine the goal of the observed agent and dynamically develop its perceived plan (Pereira et al., 2020; Shvo & Mcilraith, 2020; Vered et al., 2018). Figure 3 provides a simple illustration of the two concepts. Still, other techniques try to leverage the best of both worlds (Treger & Kaminka, 2022).

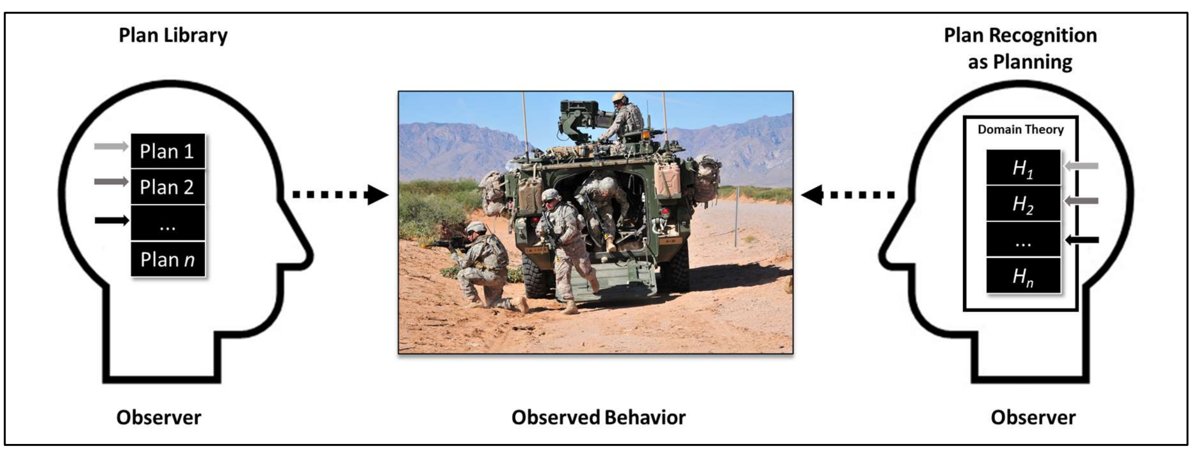

Figure 3. Plan Library vs. Plan Generation. "Stryker dismount" (US Army, 2012) is licensed under [CC BY 2.0](CC BY 2.0).

NNs and deep learning can reduce the heavy human workload needed to create plan libraries or domain expertise. However, they still require significant training to achieve their goals (Amado et al., 2018). CNNs show promise in classifying human activity and intent but struggle with situations that do not match their training (Dwivedi et al., 2019). Model-based DRL like that found in MuZero (Schrittwieser et al., 2020) could enable agents to plan their actions concerning the actions of other agents if the other agent's actions can be locally modeled. Another approach could be through few-shot learning, leading to goal recognition solutions with reduced data requirements (Dwivedi et al., 2019; Xian et al., 2020).

Graph Neural Networks can aid agents in predicting one another's following action (Liu et al., 2021). MARL enables agents to coordinate their efforts to accomplish a common goal if provided enough training (Baker et al., 2019; Ustun et al., 2020). *Multi-Agent PPO* (*MAPPO*) is an on-policy MARL algorithm for multi-agent tasks capable of handling discrete action spaces, cooperative relationships, and homogeneous agents (Yu et al., 2021). Combining the concept of intent recognition and multi-agent coordination, researchers developed a multi-agent learning algorithm that enabled agents to infer the intent of one another, allowing them to collaborate in a discrete game environment (Yuan et al., 2021).

Indirect methods such as actor-critic (Christianos et al., 2020; Ustun et al., 2020) or evolutionary learning transfer (Hou et al., 2017) could lead to agent-coordinated action. In *multi-agent deep deterministic policy gradient* (*MADDPG*), the actors operate off local information and a local policy. Still, the critic(s) who maintains a global value function governs the actor(s) (Lowe et al., 2017). Through the rewarding process, this value function can indirectly coordinate the actions of all their associated actors.

DISCUSSION

Adaptable automated forces have many challenges, including the automation of decision-making, plan recognition, and agent coordination. Fortunately, recent advancements in ML are moving modeling and simulation technology towards making adaptable OPFOR a reality.

Deep learning policy gradient algorithms are the most robust and adaptable ML promise among the decision-making approaches surveyed in this work. These approaches, such as TRPO and PPO, generate agents capable of making intelligent decisions in complex scenarios and adapting their policies in near real-time while not requiring policy retraining and the retention of complete training data to update their policies. Additionally, these approaches minimize variance in agent actions from exercise to exercise by preventing agents from straying too far from previous policy iterations. Even though a drawback to these approaches is that they require significant training data and time to be effective, these deep learning approaches provide the best agent decision-making means for automating OPFOR.

In warfare, correctly recognizing an opposing force's plan and intention allows the agent to adjust its current plan and outmaneuver the enemy. By recognizing a plan, the decision-making agent can analyze the battlefield's situation, evaluate countermeasures, and select the best response. *Plan recognition as planning* with deep learning, such as recurrent and convolution NNs, are among the best performing techniques to identify situational and action patterns. Future adaptive agents that leverage these techniques will have a combative advantage in classifying the opposition's plans, activities, and intent. Similarly, recognizing the plans and intentions of teammates could contribute to improved agent coordination (Yuan et al., 2021). Like the deep learning policy gradient approaches for adaptable decision-making, these deep learning plan recognition approaches require vast training data to fully capture the variety of possible plans, intents, and actions in military scenarios.

Lastly, implementing adaptive OPFOR in Army simulations requires standardization across simulation platforms. Agent and simulation data standards are necessary for these synthetic environments to provide agent ML approaches access to real-time simulation information and after-action review data. Through this data, agents can interpret the state of the simulation (observe), evaluate their past decisions (assess), and update their decision-making capabilities (adapt). The Army needs standardized methods and data formats for anonymized unit replays from past sim-supported training events. Only then can this data train agents capable of replacing human sim-operators. Also, current and future Army simulations need to support the integration of ML uniformly. They must provide ML algorithms with helpful information such as observation states, rewards, and action space to enable learning and agent adaptation. *OpenAI's Gym* (Brockman et al., 2016) is an example of an ML format becoming the de facto standard for developing intelligent agents. The military is creating environments that support the research of ML and DRL in support of military requirements (Brawner et al., 2022; Freeman et al., 2019; Hung et al., 2022; Liu et al., 2021; Ustun et al., 2020) but are useless to current and future Army simulations if the simulations are not modified or developed with ML in mind.

Agents that learn how to implement military tactics in an operational scenario can provide Army training audiences with a unique, realistic, and challenging OPFOR each time they train in a simulation. Current deep learning research makes it possible for agents to make intelligent decisions based on experience, identify ways to recognize and plan against modeled opponents, and coordinate agent actions. As the state-of-the-art in deep learning continues to evolve, methods become more generalized and require less data. It is critical that the Army maintains oversight of DRL advancements and facilitates its rapid integration into future military simulations.

CONCLUSION:

Threats to national security develop rapidly in the current operational environment. Army readiness ensures that the nation has trained and ready units to deploy at a moment's notice. Teams gain proficiency in their mission essential tasks through repetitive, realistic training. Deep learning agents can learn to adapt to Army training audiences. Continued research in agent decision-making, plan recognition, and coordination is the path to developing agents capable of real-time adaptation to Army training audiences. While adaptive agents continue to evolve with research, the Army must standardize unit training data formats to prepare for their future use in training any potential agents. Further, the Army must standardize current and future training simulations to facilitate the integration of these future solutions. The Army and the research community have challenges to overcome to realize fully automated, adaptive OPFOR. Still, the potential to converge on a solution and improve the readiness of Army forces is just on the horizon.

REFERENCES

Abdellaoui, N., Taylor, A., & Parkinson, G. (2009). Comparative analysis of computer generated forces' artificial intelligence. *Proceedings of the NATO RTO Modelling and Simulation Group Symposium*. https://apps.dtic.mil/sti/pdfs/ADA567877.pdf

Amado, L., Aires, J. P., Pereira, R. F., Magnaguagno, M. C., Granada, R., & Meneguzzi, F. (2018). *LSTM-based goal recognition in latent space*. arXiv. https://doi.org/10.48550/ARXIV.1808.05249

Army Futures Command Futures and Concepts Center. (2020). *Army futures command concept for maneuver in multi-domain operations 2028*. (AFC Pamphlet 71-20-1).

Avrahami-Zilberbrand, D., & Kaminka, G. A. (2014). Chapter 4 - Keyhole adversarial plan recognition for recognition of suspicious and anomalous behavior. In G. Sukthankar, C. Geib, H. H. Bui, D. V. Pynadath, & R. P. Goldman (Eds.), *Plan, activity, and intent recognition* (pp. 87-119). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-398532-3.00004-X

Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, D., & Blundell, C. (2020). Agent57: Outperforming the Atari human benchmark. In H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (pp. 507-517). PMLR. https://proceedings.mlr.press/v119/badia20a.html

Badia, A. P., Sprechmann, P., Vitvitskyi, A., Guo, D., Piot, B., Kapturowski, S., Tieleman, O., Arjovsky, M., Pritzel, A., & Bolt, A. (2020). *Never give up: Learning directed exploration strategies*. arXiv. https://arxiv.org/abs/2002.06038

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., & Mordatch, I. (2019). *Emergent tool use from multi-agent autocurricula*. arXiv. https://arxiv.org/abs/1909.07528

Ballanco, E. (2019, January 16 2019). *We need an AI-based enemy analysis tool...now!* Retrieved April 22 from https://warroom.armywarcollege.edu/articles/enemy-analysis-tool-now/

Bohemia Interactive Simulations. (2020). *VBS® control behavior pack 1*. Bohemia Interactive Simulations. https://bisimulations.com/sites/default/files/data_sheets/bisim_product_flyers_june2020_vbscontrol_pack.pdf

Brawner, K., Ballinger, C., & Sottilare, R. (2022). *Evaluating the Effectiveness of Artificially Intelligent Agents* [Submitted]. Florida AI Research Society.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). *OpenAI Gym*. arXiv. https://arxiv.org/pdf/1606.01540.pdf

Chen, S. (2017). *Learning player behavior models to enable cooperative planning for non-player characters* [Doctoral Dissertation, Carnegie Mellon University, Pittsburgh, PA].

Christianos, F., Schäfer, L., & Albrecht, S. (2020). Shared experience actor-critic for multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, *33*, 10707-10717.

Department of Defense. (2022). *Fact Sheet: 2022 National Defense Strategy*. Retrieved from https://media.defense.gov/2022/Mar/28/2002964702/-1/-1/1/NDS-FACT-SHEET.PDF

Dwivedi, S. K., Gupta, V., Mitra, R., Ahmed, S., & Jain, A. (2019). Protogan: Towards few shot learning for action recognition. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* IEEE Computer Society. https://doi.org/10.1109/ICCVW.2019.00166

Fossaceca, J., & Young, S. (2018). Artificial intelligence and machine learning for future army applications. *Proceedings Volume 10635, Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX,* SPIE Defense + Security, SPIE. https://doi.org/10.1117/12.2307753

François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, *11*(3-4), 219-354. https://doi.org/10.1561/2200000071

Freeman, J., Watz, E., & Bennett, W. (2019). Adaptive Agents for Adaptive Tactical Training: The State of the Art and Emerging Requirements. In R. Sottilare & J. Schwarz (Eds.), *Adaptive Instructional Systems* (Vol. 11597, pp. 493-504). Springer. https://doi.org/10.1007/978-3-030-22341-0_39

Grondman, I., Busoniu, L., Lopes, G. A. D., & Babuska, R. (2012). A survey of actor-critic reinforcement learning: Standard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *42*(6), 1291-1307. https://doi.org/10.1109/tsmcc.2012.2218595

Headquarters, Department of the Army. (2016). *Infantry platoon and squad*. (ATP 3-21.8).

Headquarters, Department of the Army. (2021). *Training*. (FM 7-0).

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hou, Y., Ong, Y.-S., Feng, L., & Zurada, J. M. (2017). An evolutionary transfer reinforcement learning framework for multiagent systems. *IEEE Transactions on Evolutionary Computation*, *21*(4), 601-615.

Hung, C. P., Hare, J. Z., Rinderspacher, B. C., Peregrim, W., Kase, S., Su, S., Raglin, A., & Richardson, J. T. (2022). ARL Battlespace: a platform for developing novel AI for complex adversarial reasoning in MDO. *Proceedings Volume 12113, Artificial Intelligence and Machine Learning for Multi-Domain Operations*

Applications IV, SPIE Defense + Commercial Sensing, (pp. 294-304), SPIE. https://doi.org/10.1117/12.2622227

Joint Chiefs of Staff. (2015). *Joint training manual for the armed forces of the United States*. (CJSCSM 3500.03E).

Kahn, J., Lee, A., & Hannun, A. (2020). Self-Training for end-to-end speech recognition. *Proceedings of the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 7084-7088. https://doi.org/10.1109/icassp40776.2020.9054295

Kapturowski, S., Ostrovski, G., Quan, J., Munos, R., & Dabney, W. (2019). *Recurrent experience replay in distributed reinforcement learning* [Paper]. International Conference on Learning Representations, New Orleans, LA. https://openreview.net/forum?id=r1lyTjAqYX

Kimmons, S. (2020). *Army looks to better attract gaming industry for training simulations*. Retrieved May 14 from https://www.army.mil/article/237223/army_looks_to_better_attract_gaming_industry_for_training_simulations

Lapan, M. (2020). *Deep reinforcement learning hands-on* (Second ed.). Packt publishing.

Lee, M., Valisetty, R., Breuer, A., Kirk, K., Panneton, B., & Brown, S. (2018). *Current and future applications of machine learning for the US Army*. (ARL-TR-8345). Baltimore, MD: Parsons

Liu, L., Gurney, N., Mccullough, K., & Ustun, V. (2021, 2021-12-12). Graph Neural Network based behavior prediction to support multi-agent reinforcement learning in military training simulations. *Proceedings of the 2021 Winter Simulation Conference (WSC)* IEEE. https://doi.org/10.1109/wsc52266.2021.9715433

Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in Neural Information Processing Systems*, *30*.

Millington, I., & Funge, J. (2009). *Artifcial intelligence for games* (2 ed.). CRC Press.

Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. In B. Maria Florina & Q. W. Kilian (Eds.), *Proceedings of The 33rd International Conference on Machine Learning* (Vol. 48, pp. 1928-1937). PMLR. https://proceedings.mlr.press/v48/mniha16.html

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533. https://doi.org/10.1038/nature14236

Ni, T., Li, H., Agrawal, S., Raja, S., Jia, F., Gui, Y., Hughes, D., Lewis, M., & Sycara, K. (2021). *Adaptive agent architecture for real-time human-agent teaming*. arXiv. https://doi.org/10.48550/ARXIV.2103.04439

Orkin, J. (2006). *Three states and a plan: the AI of FEAR* [Paper]. Game developers conference, San Jose, California. http://alumni.media.mit.edu/~jorkin/gdc2006_orkin_jeff_fear.pdf

Oswalt, I., Cooley, T., & Oxford, M. (2019). *Simulation based training's incorporation of machine learning* [Paper]. MODSIM WORLD, Norfolk, VA. http://modsimworld.org/papers/2019/MODSIM_2019_paper_45.pdf

Padilla, J. J. (2012). Annex 2: Military simulation systems. In *Engineering Principles of Combat Modeling and Distributed Simulation* (pp. 851-868). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118180310.oth2

Pearce, T., & Zhu, J. (2021). *Counter-Strike deathmatch with large-scale behavioural cloning*. arXiv. https://arxiv.org/abs/2104.04258

Pereira, R. F., Oren, N., & Meneguzzi, F. (2020). Landmark-based approaches for goal recognition as planning. *Artificial Intelligence*, *279*, 103217. https://doi.org/10.1016/j.artint.2019.103217

Priya Narayanan, M. V., Song Park, Anne Logie, Nick Waytowich, Mark Mittrick, John Richardson, Derrik Asher, Alexander Kott. (2021). *First-year report of ARL director's strategic initiative (FT20-23): artificial intelligence (AI) for command and control (C2) of Multi-Domain Operations (MDO)*. (ARL-TR-9192). Retrieved from https://apps.dtic.mil/sti/pdfs/AD1132186.pdf

Program Executive Office for Simulation Training and Instrumentation. *One Semi-Automated Forces (OneSAF)*. Retrieved April 15 from https://peostri.army.mil/onesaf

Program Executive Officer for Simulation, T. a. I. *Engagement skills trainer (EST)*. Retrieved June 21 from https://peostri.army.mil/engagement-skills-trainer-est

Program Executive Officer for Simulation, T. a. I. *Integrated Visual Augmentation System (IVAS) Squad Immersive Virtual Trainer (SiVT) - PEO STRI*. Retrieved May 16 from https://www.peostri.army.mil/integrated-visual-augmentation-system-ivas-squad-immersive-virtual-trainer-sivt-#:~:text=Squad%20Immersive%20Virtual%20Trainer%20%28SiVT%29%20Capability%20will%20allow,provide%20a%20realistic%20simulation%20of%20the%20operational%20environment

Roessingh, J. J., Toubman, A., van Oijen, J., Poppinga, G., Hou, M., & Luotsinen, L. (2017). Machine learning techniques for autonomous agents in military simulations—Multum in Parvo. *Proceedings of 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3445-3450. https://doi.org/10.1109/SMC.2017.8123163

Rozman, J. (2020). The Synthetic Training Environment. *Spotlight 20-6*. https://www.ausa.org/publications/synthetic-training-environment

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T., & Silver, D. (2020). Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature*, *588*(7839), 604-609. https://doi.org/10.1038/s41586-020-03051-4

Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust region policy optimization. In F. Bach & D. Blei (Eds.), *Proceedings of the 32nd International Conference on Machine Learning* (pp. 1889-1897). PMLR. https://proceedings.mlr.press/v37/schulman15.html

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms*. arXiv. https://doi.org/10.48550/ARXIV.1707.06347

Shen, Z.-Z., Yu, R., & Chen, Y.-Y. (2021). Rules-PPO-QMIX: Multi-agent reinforcement learning with mixed rules for large scene tasks. *2021 China Automation Congress (CAC)* (pp. 6731-6736), IEEE. https://doi.org/10.1109/CAC53003.2021.9728241

Shvo, M., & Mcilraith, S. A. (2020, 2020-04-03). Active goal recognition. *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 9957-9966), Association for the Advancement of Artificial Intelligence (AAAI). https://doi.org/10.1609/aaai.v34i06.6551

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., & Graepel, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140-1144. https://doi.org/10.1126/science.aar6404

Thompson, T. (2019). *The AI of Horizon Zero Dawn | Part1: Rise of the machines* [Online Video]. YouTube. https://www.youtube.com/watch?v=XxuSFBVQULY&t=335s

Toghiani-Rizi, B., Kamrani, F., Luotsinen, L. J., & Gisslen, L. (2017, 2017-10-01). Evaluating deep reinforcement learning for computer generated forces in ground combat simulation. *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 3433-3438), IEEE. https://doi.org/10.1109/smc.2017.8123161

Treger, S., & Kaminka, G. A. (2022). Towards computational modeling of human goal recognition. *Frontiers in Artificial Intelligence*, *4*, Article 737327. https://doi.org/10.3389/frai.2021.737327

United States Army Acquisition Support Center. (2022). *AVIATION COMBINED ARMS TACTICAL TRAINER (AVCATT)*. Retrieved April 22 from https://asc.army.mil/web/portfolio-item/peo-stri-aviation-combined-arms-tactical-trainer-avcatt/

United States Army Combined Arms Center. (2005). *Memorandum of agreement between the director of the TRADOC project office for ONESAF, the deputy chief of staff of operations and training, the deputy for acquisition and systems management, and the director for requirements integration* [Memorandum].

United States Army Combined Arms Center. (2018). *Synthetic Training Environment (STE) statement of need*.

United States Army Combined Arms Center. (2019). *Common Synthetic Environment (CSE) statement of need*.

US Army. (2012). *Stryker dismount* [photo]. https://www.flickr.com/photos/soldiersmediacenter/8124651746/

Ustun, V., Kumar, R., Reilly, A., Sajjadi, S., & Miller, A. (2020, November 30 - December 4). Adaptive synthetic characters for military training. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Van-Horenbeke, F. A., & Peer, A. (2021). Activity, plan, and goal recognition: A review [Review]. *Frontiers in Robotics and AI*, *8*, Article 643010. https://doi.org/10.3389/frobt.2021.643010

Van Den Bosch, K., Blankendaal, R., Boonekamp, R., & Schoonderwoerd, T. (2020). Adaptive Agents for Fit-for-Purpose Training. In *HCI International 2020 – Late Breaking Papers: Cognition, Learning and Games* (pp. 586-604). Springer International Publishing. https://doi.org/10.1007/978-3-030-60128-7_43

Vered, M., Pereira, R. F., Kaminka, G., & Meneguzzi, F. R. (2018). Towards online goal recognition combining goal mirroring and landmarks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 2112-2114). AAMAS. https://dl.acm.org/doi/abs/10.5555/3237383.3238089

Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., . . . Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, *575*(7782), 350-354. https://doi.org/10.1038/s41586-019-1724-z

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3-4), 279-292. https://doi.org/10.1007/bf00992698

Xian, Y., Korbar, B., Douze, M., Schiele, B., Akata, Z., & Torresani, L. (2020). Generalized many-way few-shot video classification. In A. Bartoli & A. Fusiello (Eds.), *Computer Vision – ECCV 2020 Workshops* (Vol. 12540, pp. 111-127). Springer. https://doi.org/10.1007/978-3-030-65414-6_10

Yu, C., Velu, A., Vinitsky, E., Wang, Y., Bayen, A., & Wu, Y. (2021). *The surprising effectiveness of PPO in cooperative, multi-agent games*. arXiv. https://arxiv.org/abs/2103.01955

Yuan, L., Fu, Z., Zhou, L., Yang, K., & Zhu, S.-C. (2021). *Emergence of theory of mind collaboration in multiagent systems*. arXiv. https://doi.org/10.48550/ARXIV.2110.00121