Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 24, 2025
*A Regional Conference of the Society for Industrial and Systems Engineering*

# Assessing the "Goodness" of Agent Behavior in Combat Simulation Models: A Framework for Evaluating the Potential of Reinforcement Learning

**Sidoni Erickson, Bedford Fagan, Jonathan Horvath,
Casey Reynolds, Eli Tate, and Matthew Dabkowski**

Department of Systems Engineering, United States Military Academy, West Point, New York 10996

Corresponding author's Email: bedford.w.fagan.mil@army.mil

**Author Note:** The first five authors executed the research described in this paper under the advisement of Colonel Matthew Dabkowski. Each member made significant contributions to this project, and a collective writing effort produced the paper. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

**Abstract:** The Research and Analysis Center (TRAC) supports Army Futures Command with timely, credible analysis to inform decisions impacting tomorrow's force. Among its most prominent tools is a suite of constructive combat simulation models allowing new capabilities, organizational structures, and tactics to be evaluated before procurement or implementation. A current initiative explores the use of reinforcement learning (RL) to train agents within these simulations. While early applications have shown promise, they have been limited to simple scenarios, leading to skepticism about their potential in more complex environments. This paper proposes a framework for assessing the "goodness" of agent behavior, creating quantifiable measures of effectiveness for assessing the benefits of RL-trained agents. Initial results using output from a well-known combat simulation suggest the approach is promising. Future work will focus on integrating relevant metrics into TRAC's accredited combat simulations to explore RL's potential in larger, more intricate combat scenarios.

*Keywords:* Reinforcement Learning, Combat Simulation, Organizational Resistance, Combat Simulation Metrics

## 1. Introduction

The Research and Analysis Center in Monterey, CA (TRAC-MTRY) is a data science research and innovation office focused on conducting applied research to improve military operations analysis. TRAC-MTRY is currently developing methods to train combat simulation entity behaviors using reinforcement learning (RL), and it has seen initial success in creating realistic behaviors for a small number of simulation entities. Despite this positive progress, TRAC's analysts may hesitate to use the RL-trained entities in a study due to the idea's infancy and their lack of exposure to the technique. The first five authors formed a capstone team charged with creating a framework to evaluate the potential of RL in combat simulation. Specifically, this capstone aims to create a testing paradigm to evaluate the "goodness" of RL-trained entities in TRAC's current simulation models, which should illuminate the pros and cons of the RL approach and reduce analyst angst. This paper summarizes this effort through a literature review of RL and combat simulation, the development of use cases and metrics, examples of assessing the realism of entity behaviors, and guidelines on overcoming organizational resistance to the integration of RL in combat simulation modeling.

### 1.1 Reinforcement Learning

RL is a form of artificial intelligence (AI) designed to mimic human behavior through a computational process. This topic is highly relevant to the Department of Defense's strategic objective to use "machines to perform tasks that normally require human intelligence" (Allen, 2020). Advancements within AI and RL have enabled models to exceed human performance in various video game formats, showing the potential value of AI (Finley, 2023). Unlike other forms of machine learning, RL learns by itself through the accumulation of rewards or punishments for its completed actions. The algorithm is not told how to complete the task, but through numerous iterations, it learns how to best complete the task at hand.

Sutton and Barto (1998) describe RL's reward function as "objective feedback from the environment" (p. 2). RL can be used to maximize rewards and explore different solutions as agents react to the environment to maximize payoffs. An agent makes a choice or decision by planning, anticipating repercussions or second-order effects, and receiving immediate feedback on the desirability of a particular position and move (Sutton & Barto, 1998). In this way, RL seeks to have agents gather data and improve performance based on a trial-and-error interaction with the environment (Allen, 2020). Agents trained with RL

algorithms can also respect ethical considerations and behavioral constraints by specifying rewards based on potential actions (Buckner, 2023).

Despite its promise, there are many challenges involved in using RL effectively in real-world problems. The first problem is the high cost associated with procuring or gaining access to the hardware required to run machine learning algorithms. The next issue, which is typically the largest one, is acquiring sufficient data to train machine learning models (Allen, 2020). The performance of the system and the quality of the output of machine learning algorithms are directly tied to the quality, depth, and breadth of the input data. The final challenge confronting the utilization of machine learning is an organization's acceptance and trust in the technology, which is exacerbated by the shortage of personnel with experience in RL. Like every piece of new technology, some people will be skeptical of its promise or resistant to its use. Older, more established organizations may be especially set in their ways and reluctant to adopt novel methods. Often this skepticism is rooted in a lack of knowledge and understanding of the new approach or item.

### 1.2 Combat Simulations

With the emergence of computer systems in the 1990s, militaries worldwide began to embrace combat simulations. These simulations and war games model the outcomes of decisions made in combat situations and the impact of technology, offering a realistic environment for users to navigate and make strategic choices. As decisions are made, the user can observe the consequences of their actions, allowing Soldiers to practice tactical and strategic skills in a controlled setting with reduced monetary costs and minimal physical risks. This has also enabled the military to refine its doctrine and standard operating procedures by testing new strategies before implementing them in the field.

In recent years, combat simulations have evolved rapidly. As technology has advanced, simulation capabilities have expanded, making these models increasingly realistic. The U.S. Army has integrated simulations into its training regimens, requiring Soldiers to pass simulation-based tests before live-fire exercises. While live drills and other traditional training methods were previously the norms, simulations have transformed training into a more practical and applicable experience, better-preparing Soldiers for real-world scenarios (Mittal & Fenn, 2024). Recently, the Department of Defense has invested in the use of AI to develop combat simulations (Finley, 2023).

TRAC aims to "conduct relevant and credible applied research to improve military operation analysis" (Thompson, 2022). One of the ways it achieves this is through constructive combat simulations, where the outputs of simulated battles are replicated and compared to real-world scenarios. This approach allows TRAC to experiment with and test innovative ideas, gaining deeper insights into the operational benefits of material, doctrinal, and organizational solutions. This includes assessing the operational effectiveness of new capabilities and tactics.

TRAC analyzes data using advanced data science and machine learning models, which enables it to produce detailed evaluations. Its work is crucial, as it helps guide decisions regarding new technologies and tactical possibilities. As part of the United States Army Futures Command, TRAC leverages this information and its ability to experiment, contributing to the development of cutting-edge tactics and strategic planning for the military (Thompson, 2022). Three of the main combat simulations used at TRAC are COMBATXXI, AWARS, and OneSAF. These simulations have been identified for the potential integration of RL (Alt, 2012).

## 2. Methodology

To implement RL-trained agents into combat simulation, four specific use cases were identified. The first use case plans to use RL-trained agents to increase the realism of combat simulations. For this use case, RL agents hold the potential to produce agent behaviors that are self-correcting and more indicative of human behavior. If successfully implemented, this would be an upgrade over the current state of having subject matter experts (SMEs) describe how agents would react across a variety of real-world scenarios, followed by analysts programming these actions into the simulation.

The second use case is using RL-trained agents to outperform conventional tactics and find new ways to fight. Currently, agents are built, and actions chosen, by SMEs based on Army doctrine or likely future modifications of it. RL-trained agents could drive doctrinal changes by loosening the constraints on their actions, thereby increasing their tactical flexibility when fighting the enemy. By using RL-trained agents, SMEs can shift their focus from prescribing the simulation's input to evaluating the simulation's output for unorthodox yet effective tactics.

The third use case would use RL-trained agents to test revolutionary equipment and evaluate novel capabilities. Testing these futuristic capabilities can expose hidden capability gaps and inform future requirements on the evolution of tactics and equipment. Presently, programming these new technologies into combat simulations is a labor-intensive process. In an RL-enabled combat simulation, agents could evaluate new technology by using an "arms room" approach, where agents iteratively

select weapons with different capabilities to maximize their rewards. The output of such exploration could then be used to identify what capabilities might make an outsized impact on tomorrow's battlefield.

The final use case is using RL-trained agents to expose flaws in TRAC's current combat simulation software. Currently, SMEs create simulations in hopes of replicating complex real-world scenarios, but they may fail to identify hidden flaws in the simulated environment. These errors could manifest in many ways and compound over a simulation run, negatively impacting the performance and realism of TRAC's experimentation. By allowing RL-trained agents unconstrained freedom of operation, TRAC's analysts could identify and correct these flaws.

Figure 1 summarizes the four use cases into an evaluation hierarchy that frames each use case's potential as a question relative to the status quo. For example, the first use case – *increase the realism of agent behavior* – aims to answer the question: [When compared to TRAC's current combat simulations], can RL-trained agents better emulate how Soldiers behave on the battlefield? If the answer to this question is "Yes," then RL-trained agents have goodness vis-à-vis TRAC's charter "to improve military operation analysis" (Thompson, 2022).

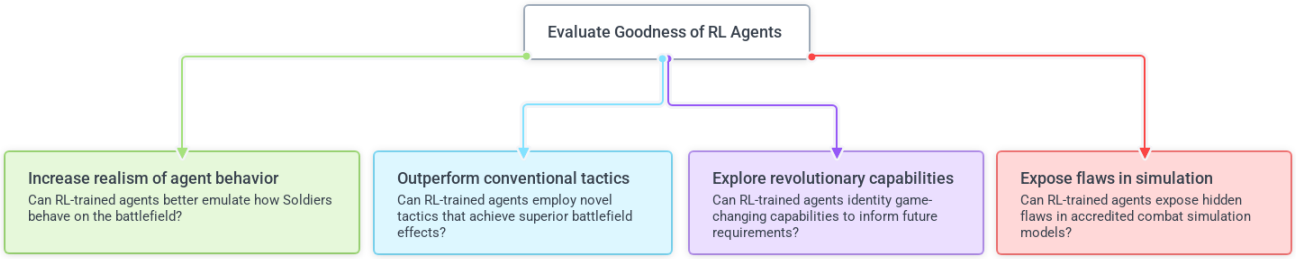


Figure 1. Topmost-Level Evaluation Hierarchy of RL-Trained Agents in Combat Simulation

## 3. Metrics

In the sections that follow, we will focus on *Use Case 1 – Realism*. As seen in Figure 2, our method expands the evaluation framework centered on the realism of soldier behavior. Within this use case, realism is evaluated by how soldiers should fight versus how soldiers actually do fight. More specifically, we will look into how soldiers actually fight by evaluating how they conserve ammunition, deviate from the plan or standard operating procedures (SOPs), and preserve life by helping themselves or their fellow soldiers when they become wounded. Additionally, individual soldier actions contribute to realism, such as a soldier reloading his magazine after an engagement.

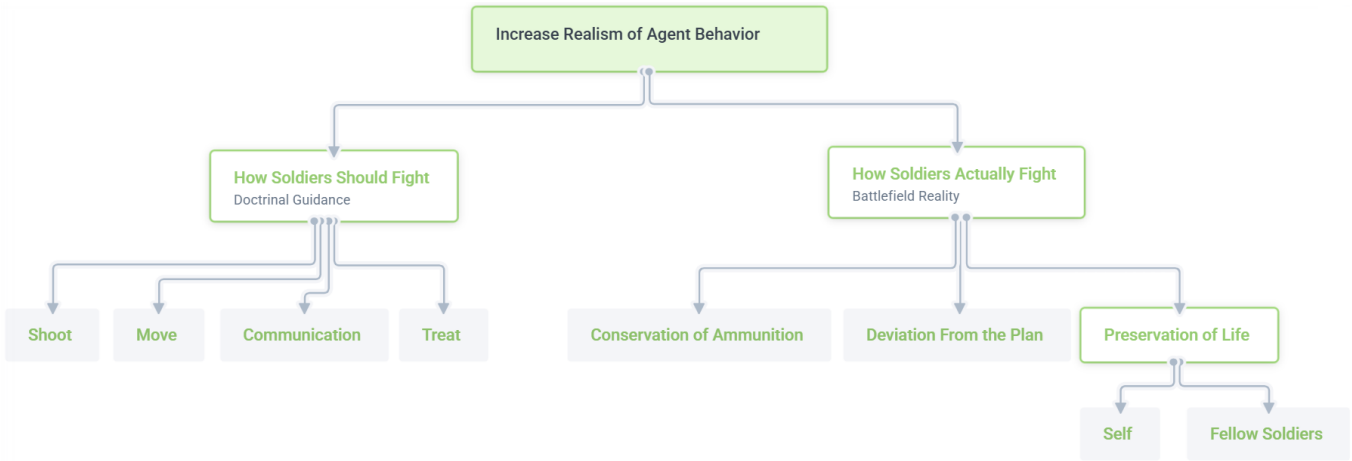


Figure 2. Realism Evaluation Hierarchy

ISBN: 97819384969-9-8

### 3.1. Metrics Applied to Simple IWARS Scenario

To assess these metrics, we applied the evaluation criteria to a scenario we developed within the Infantry Warrior Simulation (IWARS), which is a "Soldier-centric modeling capability to conduct integrated, multi-domain analyses that explore the complex relationships between [s]oldiers, their equipment, and their battlefield environment" (Samaloty et al., p. 29). In the scenario, nine U.S. Army personnel come into contact with an enemy observation post (OP) manned by a two-soldier buddy team. An overhead view of this IWARS simulation is shown in Figure 3, where blue dots represent U.S. soldiers, red dots denote enemy soldiers, and orange lines indicate prescribed routes for maneuver.

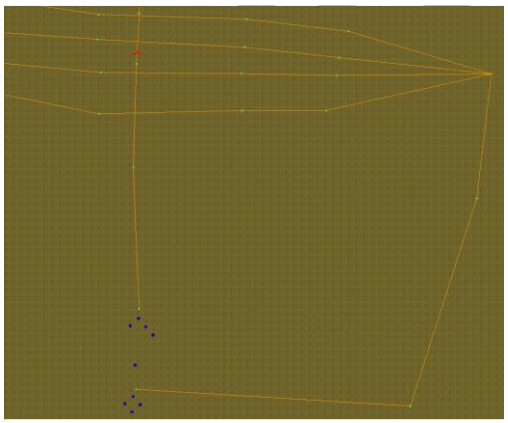

Figure 3. Simple IWARS Simulation

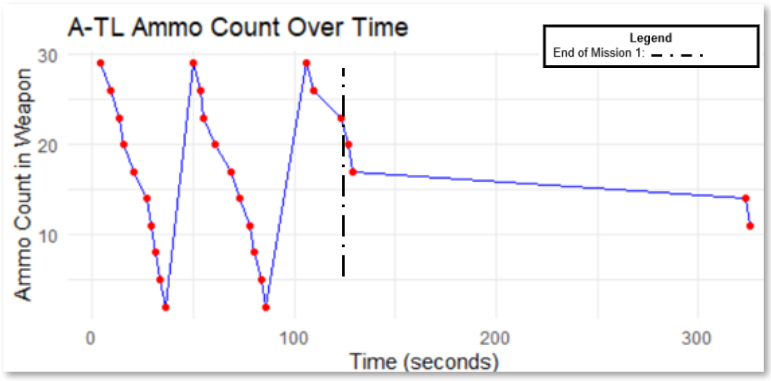


Figure 4. Alpha Team Leader Ammunition Count

The U.S. squad has a variety of weapons organic to its structure. One of the weapons is an M4 assault rifle with a magazine capacity of 30 rounds. The squad is broken into two, four-man teams controlled by a squad leader. Once in contact, the squad reacts by conducting *Battle Drill 2A - Squad Attack* (Department of the Army, 2020). After the squad clears through the enemy OP, it continues its mission to a follow-on objective with two more enemy soldiers. To evaluate the realism of agent actions, we looked at ammo expenditure, survivability, and adherence to standard operating procedures (SOPs) over 50 independent simulation runs, where the end of Mission 1 was defined as the death of both enemy soldiers in the OP.

Figure 4 shows the ammunition remaining in the Alpha Team Leader's (A-TL's) weapon during a specific run. As the A-TL fires his rifle, the rounds remaining in the magazine decrease. Upon the full expenditure of each magazine, the A-TL loads a new one, increasing the y-axis value to 30, and the cycle repeats, as each fresh magazine contains 30 rounds. As the point man in the base of fire element, the A-TL was in contact with the enemy for the longest amount of time, and his ammunition expenditure followed a fairly constant rate. While this would occur if the enemy OP was overwhelming the U.S. soldiers, a decreasing rate of fire is more realistic here. Known as the conversation of ammunition in Army doctrine (Department of the Army, 2018), the A-TL's ammunition expenditure rate should have slowed down as his remaining supply of ammunition dropped. Even though Figure 4 depicts one run, it highlights the agent's inability to evaluate the value of each remaining round. If an RL-trained agent's ammunition-remaining count function exhibited a concave-up shape, it would suggest the agent had learned to conserve its ammunition.

Similar to Figure 4, Figure 5 graphically captures another element of realism – how soldiers might deviate from the pre-mission tactical plan based on during-mission tactical developments. Specifically, U.S. doctrine dictates that an attacking force's soldiers should outnumber a defending force's soldiers by at least 3 to 1 (Department of the Army, 2018). In the IWARS scenario discussed above, nine U.S. infantrymen are assaulting two enemy soldiers. The attacking-to-defending ratio is 4.5 to 1, and the initial decision to assault is doctrinally sound. However, as Figure 5 shows, the attacking squad sustained three casualties during the first 150 seconds of the first engagement, dropping its combat-effective strength to six. Moreover, these losses would have cascading effects, as combat-effective members of the squad would provide security and medical treatment for their wounded comrades. This would undoubtedly drive the squad's effective strength below the requisite ratio of 3 to 1, causing the squad to abort their follow-on mission to attack the second OP. Figure 6 shows a histogram of the number of combat-effective U.S. soldiers remaining after Mission 1 for the 50 runs. With a mean of 7.36, if we assume the casualties are normally distributed, in roughly 16% of instances, the U.S. force should not have continued to the second engagement with 6

or fewer combat-effective soldiers. Using RL, simulated agents could be trained to recognize this ratio and determine when to continue or abort the attack, increasing the realism of the squad's actions without the need for SME input or coded instructions.

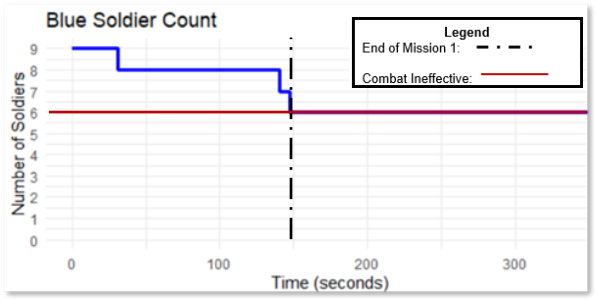


Figure 5. Combat-Effective U.S. Soldiers over time during the First Mission (from the run depicted in Figure 4)

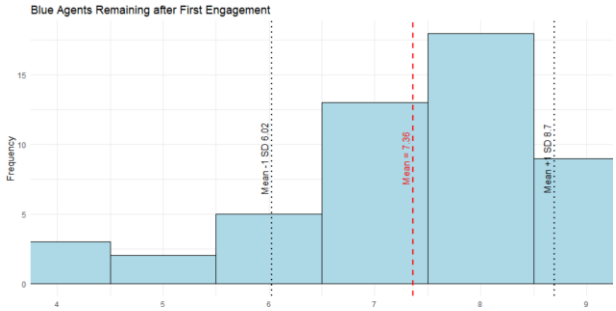


Figure 6. Combat-Effective U.S. Soldiers remaining after the First Mission (from all 50 runs)

### 3.2. Future Work

Due to time and resource constraints, we were unable to explore all possible metrics for evaluating realism. However, we did develop other possible metrics that could be beneficial in the evaluation of RL agents. These metrics, along with their significance and potential assessment process, are listed in Table 1 below. These metrics would enhance the evaluation of agents by providing additional data points on the realism of their behavior within the simulation environment.

Table 1. Potential Metrics for Evaluating the Realism of RL-Trained Agents in Combat Simulations

| Potential Metric | Significance (In the real world…) | Assessment |
|---|---|---|
| Distance between soldiers | Soldiers maintain a distance from their teammates that maximizes protection from enemy fire while ensuring cohesion and communication. | Utilize the location data for each simulated soldier to calculate the distance between them and compare these distances to doctrinal guidance and the observed behaviors of humans. |
| Use of alternative assets (drones, fires, etc.) | Soldiers employ external and indirect fire assets to increase combat power. | Use the rates of fire and alternative assets to guide agent behavior compared to actual usage. |
| Support for incapacitated soldiers | Soldiers provide aid to and protect casualties. | Measure the number of soldiers that stay behind with casualties with treatment needs. |

## 4. Combating Organizational Resistance

There are many potential upsides for the use of RL in combat simulations but implementing RL would represent a major change for TRAC. Like any change in any industry, there will be those who are hesitant to accept it. Through research, we have found industries that overcame organizational resistance and adopted similar technological breakthroughs.

The first organization is the National Football League (NFL). The NFL is a billion-dollar industry whose coaches have traditionally made difficult game decisions based on gut instinct and prior experience. Over the past 15 years, this decision-making approach has gradually given way to advanced analytics and statistical analysis. This can be seen most notably in game situations where a team chooses whether to "go for it" on fourth down. Before the use of analytics, it was rare for a team to attempt to convert on fourth down. However, in the past decade, these attempts have more than doubled (Sutelan, 2023). The second example involves DeepSeek (2015), a Chinese startup AI company that made a major breakthrough when it decided to use RL to train and refine their models. This process allowed Deepseek's RL language models to outperform or be very competitive with similar models in the industry at a significantly cheaper cost (Yang, et al., 2025).

These two examples lend themselves to overcoming organizational resistance in TRAC in two unique ways. To begin, DeepSeek is a small company that faced extinction: they simply do not have the resources or the time to compete with the market leaders such as OpenAI and Meta. Forced to adapt, DeepSeek turned to RL to power their model and discovered unprecedented success in a short time. TRAC does not see competition directly, as it is the preeminent leader in its respective

field. However, its competition comes from near-peers, such as China and Russia, and failure to adapt or evolve in the field could be devastating in the future. Similarly, TRAC parallels the NFL in that it will need to show the efficacy of RL to remove the resistance to its adoption. Change in procedure, such as fourth-down play calling, results in initial hesitation within an organization. Through proven results, trust is built which serves as a catalyst for increased change. Organizational resistance should be overcome if TRAC is willing to take the leap and build an effective RL model.

## 5. Conclusion

The use of RL agents in simulation holds significant promise in the field of combat modeling, but further work is required to successfully integrate the evaluation tools into practical applications. First, the RL capabilities of combat simulations must be improved in order to test their effectiveness. The difficulty of integrating RL agents into simulations hinders their evaluation. Next, further research is required to refine quantitative evaluation metrics. As discussed in Section 3.2, there are other metrics that can be used to evaluate RL agents, and Table 1 highlighted several examples. Additionally, there may be other potential realism metrics we have yet to consider, which could further refine the evaluation criteria for RL agents. There are also opportunities for more research into the other use cases for RL agents, as seen in Figure 1. Finally, this work provides two suggestions for softening an organization's resistance to integrating RL into combat simulations, namely embracing an existential urgency to innovate (i.e., innovate or die) and thoroughly assessing the efficacy of the approach (i.e., verify whether RL actually makes the Army better). While there is still much work to do before RL agents in combat simulation can be completely assessed, our research provides insight into how future work can tackle this problem. Overall, the use of RL has innovative promise in the field of combat simulation. If successfully integrated, RL could revolutionize decision-making and strategic planning in warfare.

## 6. References

Allen, G. (2020). Understanding AI Technology - A concise, practical, and readable overview of Artificial Intelligence and Machine Learning technology designed for non-technical managers, officers, and executives. https://apps.dtic.mil/sti/pdfs/AD1099286.pdf

Alt, J. K. (2012). *Learning from Noisy and Delayed Rewards: The Value of Reinforcement Learning to Defense Modeling and Simulation*. (Publication No. ADA567384) [Doctoral Dissertation, Naval Postgraduate School]. https://apps.dtic.mil/sti/pdfs/ADA567384.pdf

Buckner, C. J. (2023). *From deep learning to rational machines: What the history of philosophy can teach us about the future of artificial intelligence*. New York: Oxford University Press. https://doi.org/10.1093/oso/9780197653302.001.0001

DeepSeek-AI, Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R.,…Zhang, Z. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. https://doi.org/10.48550/arXiv.2501.12948

Department of the Army. (2018). *Infantry Rifle Company* (ATP 3-21.10). https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN8519_ATP%203-21x10%20Final%20Web.pdf

Department of the Army. (2020). *Ranger handbook* (TC 3-21.76). https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN3039-TC_3-21.76-000-WEB-1.pdf

Finley, M. G. (2023). *Applied reinforcement learning wargaming with parallelism, cloud integration, and AI uncertainty*. (Publication No. AD1213264) [Master's Thesis, Naval Postgraduate School]. https://apps.dtic.mil/sti/trecms/pdf/AD1213264.pdf

Mittal, V., & Fenn, J. E. (2024). Using combat simulations to determine tactical responses to new technologies on the battlefield. *The Journal of Defense Modeling and Simulation*. https://doi.org/10.1177/15485129241239364

Samaloty, N. N. E., Schleper, R., Fawkes, M. A., & Muscietta, D. (2007). Infantry Warrior Simulation (IWARS): A Soldier-Centric Constructive Simulation. *Phalanx, 40*(2), pp. 29–31. https://www.jstor.org/stable/24909630

Sutelan, E. (2023). *NFL 4th Down Conversion Chart, Explained: Breaking Down the NFL's Success Rates by Distance & More to Know*. The Sporting News. https://www.sportingnews.com/us/nfl/news/nfl-fourth-down-conversion-chart-rate-by-distance/vofkeub6xwms6imajxqkfipp

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press. https://muse.jhu.edu/pub/6/oa_monograph/book/60836

Thompson, M. (2022). *TRAC-Monterey leverages state-of-the-art data science to inform future force planning*. U.S. Army. https://www.army.mil/article/258100/trac_monterey_leverages_state_of_the_art_data_science_to_inform_future_force_planning