AI for Tactical Planning: Route Generation and Adaptive Behavior

Stephen Moore
University of Central Florida
Orlando, Florida, USA
st690445@ucf.edu

Lauren Lucernoni
University of Central Florida
Orlando, Florida, USA
la685227@ucf.edu

C. Lynn Yingling
University of Central Florida
Orlando, Florida, USA
ch192703@ucf.edu

***Abstract*—Modern military operations require rapid, context-sensitive decision-making in complex environments. This study explores two computational methods to support tactical planning: an A* pathfinding algorithm enhanced with doctrinal heuristics, and a reinforcement learning system based on Proximal Policy Optimization (PPO) trained through a progressively challenging scenario. The A* approach focuses on generating routes that prioritize cover, concealment, and threat avoidance, while the PPO model builds adaptive and tactical decision-making behaviors in varied combat situations. We used a mixed-method evaluation—combining quantitative performance metrics with qualitative assessments from military subject matter experts to assess both methods. Results show that A* produces consistent, terrain-aware routes, and PPO adapts well to different mission types. Both systems were found to perform effectively within their respective domains. While each approach demonstrated measurable benefits, the study also highlights key areas for improvement—particularly in doctrinal integration and computational performance—suggesting directions for future refinement of tactical planning tools.**

Keywords—reinforcement learning, Proximal Policy Optimization, A* pathfinding, tactical planning, military AI, curriculum learning, mixed-methods evaluation, decision support systems, kernel filtering, combat simulation

I. Introduction

Modern military operations demand rapid, accurate decision-making in complex environments. Course of Action (COA) analysis and wargaming provide structured approaches to evaluate strategic options, but traditional methodologies have been criticized for their manual and time-intensive nature, particularly in dynamic operational contexts [1]. Our research addresses this critical gap by using Proximal Policy Optimization (PPO) within a Multi-Agent Reinforcement Learning (MARL) framework to improve adaptability and effectiveness in military planning [2], [3]. Future combat operations will require inexperienced personnel to assume planning roles, potentially leading to suboptimal tactical decisions [1]. While existing planning frameworks ensure consistency, they often become bottlenecks in rapidly evolving combat scenarios [1]. This challenge is amplified as operations increase in complexity and resources become constrained. Our research specifically addresses two key capability gaps: rapid generation of courses of action and 2) rapid analysis of the courses of action absent of any personal bias / inexperience.

By leveraging reinforcement learning techniques, specifically PPO, combined with the A* pathfinding algorithm for optimal route planning, we aim to enhance tactical decision-making while maintaining alignment with military doctrine [2], [6], creating a framework that generates and evaluates tactical plans more efficiently [1].

This paper presents a mixed-method research design combining quantitative analysis of route effectiveness with qualitative feedback from military planners. We evaluate our approach across three distinct tactical scenarios representing different operational challenges. Through statistical analysis and expert evaluation, we demonstrate both the capabilities and limitations of Artificial Intelligence (AI)-assisted tactical planning.

II. Background

Military decision-making has traditionally relied on structured frameworks such as the Military Decision-Making Process (MDMP) [6]. While these frameworks provide rigor and repeatability, they also rely heavily on manual input and are often time-consuming—making them less effective in dynamic or resource-constrained operational environments [1], [3].

Evolution of AI in Military Planning

AI applications in military planning have evolved from rule-based systems to more sophisticated approaches. Early efforts using decision trees and neural networks provided some utility but struggled with complex scenarios requiring adaptation. The introduction of reinforcement learning marked a significant advancement, enabling agents to learn from feedback and adapt behaviors over time [4].

Recent work by Goecks and Waytowich (2024) demonstrated how generative pre-trained transformers can accelerate Course of Action (COA) development [1]. Although generative approaches show potential, they must be carefully implemented to ensure their outputs align with military doctrine, so they do not result in tactically flawed decisions [1].

Existing machine learning methods have struggled to adapt to the high-stakes nature of military decision-making. Rule-based systems lack flexibility required to operate effectively in uncertain, complex environments [3]. More advanced models—such as deep neural networks and reinforcement learning (RL) agents—have shown promise, but they also face limitations. These include difficulty generalizing to unseen scenarios, dependence on extensive training data, and high computational demands that can restrict real-time applicability in resource-constrained settings [4], [5].

Reinforcement Learning and Tactical Path Planning

Our research builds upon two distinct approaches to tactical planning challenges. For adaptive tactical behaviors, we examine reinforcement learning's effectiveness in decision-making [4]. Proximal Policy Optimization (PPO) represents a significant advancement in this domain, offering improved sample efficiency and performance stability compared to earlier RL approaches [2]. PPO's policy gradient method allows for incremental updates that balance exploration with exploitation, making it particularly suitable for optimizing mission success and resource utilization in tactical scenarios [2], [4].

Separately, for route planning challenges, our research employs the A* pathfinding algorithm enhanced with custom tactical heuristics. These enhancements consider factors like cover, concealment, and threat exposure [7]. The algorithm's ability to efficiently search large state spaces while incorporating tactical considerations makes it well-suited for generating routes that balance computational efficiency with doctrinal soundness.

Research Gap and Contribution

Despite advances in AI for military applications, significant gaps remain in turning these technologies into practical tools for tactical planning [3]. The primary limitations we address include:

1) The disconnect between computational efficiency and tactical soundness in route generation (addressed through A*)
2) The challenge of developing systems that adapt to varied operational environments (addressed through PPO)
3) The lack of comprehensive evaluation methods that combine quantitative performance with qualitative doctrinal alignment

Our research addresses these limitations through two independent but complementary approaches. We investigate PPO reinforcement learning for developing adaptive tactical behaviors that optimize mission success, evaluated through quantitative performance indicators [2], [4]. Separately, we examine A* pathfinding with tactical heuristics for generating routes that balance computational efficiency with doctrinal soundness, evaluated through both quantitative metrics and qualitative expert assessment [7]. This mixed-methods approach for evaluating the A* algorithm provides more comprehensive insights into route planning quality than quantitative measures alone.

III. METHODOLOGY

This research investigates how different artificial intelligence techniques enhance military tactical planning across varied operational environments. Specifically, we examine: (1) how effectively an A* pathfinding algorithm with tactical enhancements generates routes balancing computational efficiency with doctrinal soundness; and (2) to what extent a Proximal Policy Optimization (PPO) reinforcement learning approach develops adaptive tactical behaviors optimizing mission success [1], [2].

We hypothesize that an A* algorithm enhanced with tactical heuristics will generate routes that maintain consistent quality across varied terrain while demonstrating adaptation based on terrain complexity [7]. However, we anticipate computational optimization alone may not fully capture doctrinal considerations necessary for tactically sound plans. Separately, we hypothesize that a PPO reinforcement learning system will develop adaptive tactical behaviors responding to different scenario parameters, demonstrating measurable differences in tactical approach when trained through a progressive curriculum [2], [4].

Our approach employs a mixed-method design, combining controlled simulations with expert qualitative feedback for the A* pathfinding component. For route planning, we employ the A* algorithm enhanced with custom heuristics prioritizing tactical considerations such as cover, concealment, and threat exposure [7]. Independently, we implement PPO within a multi-agent framework to enhance broader tactical decision-making, evaluated through quantitative performance indicators [1], [2], [4].

Unlike previous models focusing on either real-time or turn-based strategies in isolation [2], this study implements a step-based execution model. We differentiate our approach by leveraging curriculum learning techniques to refine agent behavior without extensive retraining. The methodology structures decision-making across command levels to ensure strategies align with established military doctrine. The following sections detail the specific approaches, experimental conditions, procedural design, evaluation metrics, and implementation.

A. Approach

Our research employs two distinct AI techniques to address different aspects of military tactical planning. The A* pathfinding algorithm focuses on route generation, while the PPO reinforcement learning approach develops adaptive tactical behaviors. Each technique is evaluated independently using appropriate methodologies.

A* Pathfinding for Route Planning

Our A* algorithm enhances infantry squad movement planning through a structured approach prioritizing cover, concealment, elevation, and threat exposure. Unlike standard implementations, our system refines the search area using sophisticated kernel filtering techniques. These filters slide a weighted grid over terrain data to assess locations based on their surroundings, identifying areas with better tactical advantages [8]. By ranking positions based on concealment, elevation, and enemy engagement zones, the system significantly reduces non-viable movement locations.

The technical workflow begins with terrain preprocessing and filtering. A selection process then narrows possible locations further, ranking them based on predefined criteria. This computational approach transforms what would be a prohibitively large search space into a manageable one without sacrificing tactical considerations. The final routes are generated using an A* path-planning algorithm enhanced with these filters, ensuring optimized travel efficiency while minimizing exposure to hostile forces.

This approach directly addresses our first research question by balancing computational efficiency with doctrinal soundness. The algorithm's ability to adapt routes based on terrain complexity aligns with our A* Pathfinding Hypothesis, while the need for expert evaluation acknowledges the limitation of computational optimization alone.

PPO Reinforcement Learning for Tactical Behaviors

For developing adaptive tactical behaviors, we implemented a Multi-Agent Reinforcement Learning system using Proximal Policy Optimization (PPO) with a sophisticated curriculum learning approach. This addresses our second research question by optimizing mission success through quantitative performance indicators.

The PPO training utilized a progressive curriculum with four distinct levels of increasing complexity:

1) **Level 1: Basic Navigation:** Training began on a bare map with no terrain features or enemies. This level contained nine stages with three distance categories:

- Close distance stages (1-3): Agents learned to navigate to objectives at varying locations from nearby starting positions
  - Medium distance stages (4-6): Increased travel distance with adjusted hyperparameters for more complex navigation
  - Far distance stages (7-9): Significantly increased distances requiring extended planning capabilities

2) **Level 2: Terrain Complexity:** Once agents mastered basic navigation, terrain features were introduced while maintaining the same objective and unit configurations as Level 1. This level featured:
   - Increased value coefficients (50% higher than Level 1)
   - Extended episode lengths (+500 steps per episode)
   - Higher entropy coefficients to encourage exploration in complex terrain
3) **Level 3: Enemy Introduction:** After developing terrain navigation skills, agents faced enemies positioned in front of objectives with:
   - Reduced engagement range (10 cells)
   - Significantly increased value coefficients for enemy presence
   - Six progressive stages alternating between bare and terrain maps
4) **Level 4: Advanced Combat:** The final level introduced more challenging combat scenarios with:
   - Half-normal engagement range (20 cells)
   - Enhanced value coefficients for tactical decision-making
   - Higher entropy coefficients to encourage exploration of combat tactics

To make reinforcement learning computationally feasible, we discretized the various parameters of the actions to reduce the action space from approximately 24 billion to 4.06 million state-action pairs—a 99.98% reduction—while maintaining tactical richness. Our action space encompasses five fundamental military operations:

1) MOVE (0): Basic movement operations with direction and distance parameters
2) ENGAGE (1): Direct fire at enemies with target position and fire control settings
3) SUPPRESS (2): Area suppression fire
4) BOUND (3): Tactical bounding movement
5) CHANGE_FORMATION (4): Modify unit formation types and orientation

We implemented sophisticated target validation to ensure militarily sound engagements. Targets must be within the unit's observation range, engagement range, and sectors of fire, with line of sight unobstructed by terrain. Invalid target selections are either re-routed or skipped. This approach provides a tactically authentic set of actions while keeping the learning problem computationally tractable.

Our neural network architecture processed different types of information through specialized encoder modules before combining them, mirroring how military commanders process battlefield data. The actor network used modular encoders for agent state (32 units), tactical information (16 units), and spatial information (32 units each for friendly units, enemies, and engagement sectors), with two 160-neuron hidden layers in the combined network. The critic network mirrored this structure but used deeper combined layers (256 units) for more accurate value prediction.

We employed several advanced optimization techniques to enhance learning stability and efficiency. These included advantage normalization to reduce variance in policy gradient estimates, return normalization for critic training, gradient clipping (0.5) to prevent excessively large updates, adaptive learning rate scheduling, entropy coefficient annealing to balance exploration and exploitation, and early stopping with multiple criteria beyond simple reward metrics.

This curriculum-based approach aligns directly with our PPO/MARL Hypothesis, enabling agents to develop measurably different tactical approaches when trained through varying terrain, enemy configurations, and engagement parameters.

Mixed Method Approach

To comprehensively evaluate the A* route planning algorithm, we employed a mixed-method research design—a choice that addresses our Methodological Hypothesis. This approach was selected because quantitative metrics alone cannot fully capture the tactical soundness of generated routes, while expert evaluation in isolation lacks objective performance measures.

The quantitative component analyzes route effectiveness through metrics including threat exposure, concealment levels, and cover utilization. These measurements provide objective data on the computational efficiency of the generated routes. Route quality is assessed through numerical scores that can be compared across different terrain environments to validate the consistency hypothesis.

The qualitative component involves systematic assessment by military domain experts—five U.S. Army officers with combat experience from Infantry, Field Artillery, and Armor branches. Each expert evaluated the tactical viability of the algorithm-generated routes using standardized criteria: six binary performance measures ("Go"/"No-Go") and three Likert scale ratings (1-5) for doctrinal soundness, operational feasibility, and novel tactical insight. This structured evaluation framework provides consistent assessment of tactical principles that cannot be easily quantified.

We validated the reliability of expert assessments using Cronbach's alpha coefficients, which demonstrated strong agreement among experts for established tactical criteria (Doctrinal Soundness: $\alpha = 0.750$, Operational Feasibility: $\alpha = 0.833$), while showing more diverse perspectives on tactical innovation (Novel Tactical Insight: $\alpha = 0.429$).

This mixed-method approach bridges the gap between computational metrics and real-world tactical requirements, providing more comprehensive insights than either approach alone. While the A* algorithm optimizes routes based on quantifiable parameters, the qualitative expert evaluation determines whether these routes would be viable in actual combat operations, directly testing our hypothesis that computational optimization alone cannot fully capture doctrinal considerations. Fig. 1 (Infantry Movement Planning System) depicts this blended approach which allows us to gain data-driven optimization with tactical feasibility, addressing gaps in traditional movement planning methods.

B. *Simulation Environment Validation*

Before implementing the full experimental procedures, we conducted a systematic validation of the military tactical simulation environment to ensure it accurately modeled critical tactical considerations while maintaining computational efficiency. This validation process provided the foundation for both the A* pathfinding and PPO reinforcement learning experiments.

Tactical Position Filtering Validation (A* Environment)

The tactical position filtering system for the A* pathfinding environment was tested to verify its ability to reduce non-essential data while preserving high-quality tactical positions. Position quality scores demonstrated appropriate differentiation between tactical roles, with assault positions showing higher average quality scores (0.52-0.54) compared to support positions (0.45-0.47). This alignment with operational priorities validated the position selection algorithm's ability to incorporate doctrinal considerations into its filtering process, ensuring that the A* algorithm would have high-quality candidate positions to consider during route planning.

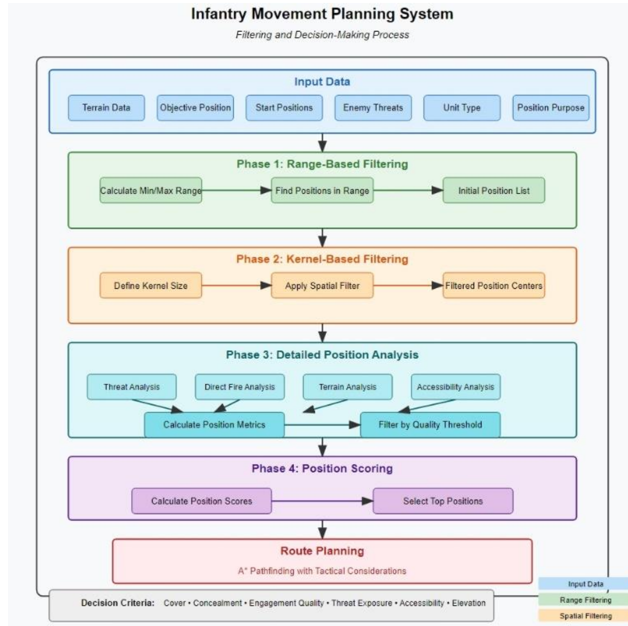

Fig. 1. Systems architecture shows the filtering and decision-making process for infantry movement planning. The workflow progresses from terrain preprocessing through kernel filtering to path generation.

Combat Mechanics Validation (PPO Environment)

We verified the PPO reinforcement learning environment's combat mechanics by comparing weapon effects against expected tactical behavior. Automatic weapons (e.g., M249 SAW) demonstrated appropriate suppression effects relative to single-shot weapons (M4 Carbine), and terrain properly influenced suppression effectiveness as expected in combat operations. Across 50 simulated engagements, hit probability calculations ($M = 0.89$, $SD = 0.05$) aligned with theoretical models. Suppression duration ($M = 11.7$ steps, $SD = 2.3$) and intensity ($M = 0.73$, $SD = 0.12$) demonstrated patterns consistent with field observations, ensuring the environment properly modeled critical combat dynamics that would influence agent learning during PPO training.

Computational Performance Validation (Both Environments)

To ensure both simulation environments would support efficient experimentation, we measured key system performance metrics. For the A* environment, the tactical filtering system processed 40,000 potential movement options down to 10 optimal positions in 4.23 seconds ($SD = 0.45$), while route planning algorithms completed within 1.87 seconds ($SD = 0.28$). For the PPO environment, engagement resolution, including line-of-sight calculations and hit probability determination, averaged 0.007 seconds ($SD = 0.002$) per instance. These metrics confirmed that both environments maintained sufficient computational efficiency to support complex tactical simulations while preserving necessary tactical fidelity.

This validation process ensured that our simulation environments maintained realistic operational parameters while providing efficient computational performance, creating a reliable foundation for our subsequent experiments with both A* pathfinding and PPO reinforcement learning approaches.

C. Condition / Scenario Development

To thoroughly evaluate both AI approaches in our research, we developed distinct testing environments and scenarios for the A* pathfinding algorithm and the PPO reinforcement learning approach. Each testing framework was designed to address the specific capabilities and limitations of its respective algorithm while maintaining tactical realism.

A* Pathfinding Testing Environment

The A* pathfinding algorithm was tested across three terrain scenarios to evaluate its ability to generate tactically sound routes with appropriate consideration for cover, concealment, and threat exposure. For each scenario, we generated nine routes (three each for Assault, Support, and Reserve positions) and recorded key metrics including threat score, cover score, concealment score, route distance, and overall quality.

The implementation combines position identification and route planning capabilities to create a comprehensive tactical movement planning system. The position identification component employs a progressive filtering process that includes range-based area reduction, spatial filtering with unit-size kernels (7×7 cells for teams, 12×12 cells for squads), pooled metrics analysis, and position scoring using purpose-specific weights.

Each tactical role has specific requirements codified in the system. Assault positions require higher minimum cover (0.4) and concealment (0.3), close range to objective, wider engagement arc (90°), and maximum threat exposure of 0.7. Support positions need moderate cover (0.3) and concealment (0.2), medium range to objective (400-800m), standard engagement arc (60°), and can tolerate a maximum threat exposure of 0.8. Reserve positions balance cover (0.3) and concealment (0.3), maintain further range from the objective, have minimal engagement requirements, and require the lowest acceptable threat exposure of 0.6.

The tactical parameter weights used for verification are outlined in Table I (Tactical Parameter Weights for Verification), carefully designed to reflect realistic battlefield conditions and align with established military doctrine. These weights directly influence the cost function used by the A* algorithm, affecting how it values different terrain features and tactical considerations. The system supports three distinct routing strategies:

1) **Concealment-Focused**: Prioritizes routes with minimal visibility to enemy positions (concealment weighted at 0.5, threat at 0.3, cover at 0.2)

2) **Cover-Focused**: Emphasizes protection from enemy fire (cover weighted at 0.5, threat at 0.3, concealment at 0.2)

3) **Balanced**: Even distribution between all factors (threat protection at 0.4, cover and concealment each at 0.3)

TABLE I. TACTICAL PARAMETER WEIGHTS FOR VERIFICATION

Tactical Parameter Weights for Verification

Customizable Parameters by Tactical Domain

Domain	Parameter	Default Value	Range	Source File
1. TERRAIN CONSIDERATIONS				
Terrain Movement Costs	terrain_costs	BARE: 1.0 SPARSE_VEG: 1.2 DENSE_VEG: 1.5 WOODS: 2.0 STRUCTURE: ∞	0.0 - ∞	WarGamingEnvironment_vTest.py
Line of Sight Degradation	terrain_degradation	BARE: 0.0 SPARSE_VEG: 0.05 DENSE_VEG: 0.15 WOODS: 0.25 STRUCTURE: 1.0	0.0 - 1.0	WarGamingEnvironment_vTest.py
Elevation Effects	elevation_observation_bonus elevation_cover_penalty	0.1 (10%)	0.0 - 1.0	WarGamingEnvironment_vTest.py
2. POSITION ANALYSIS SETTINGS				
Position Requirements	min_cover min_concealment min_engagement max_threat	SUPPORT: 0.3 ASSAULT: 0.4 RESERVE: 0.3 (varies by purpose) (varies by purpose)	0.0 - 1.0	tactical_position_analyzer.py
Position Scoring Weights	position_scoring_weights	SUPPORT: fire_quality: 0.4, cover: 0.2 concealment: 0.2, threat: -0.2 ASSAULT/RESERVE: (different weights) (see detailed documentation)	-1.0 to 1.0	tactical_position_analyzer.py
Position Bonuses	accessibility_importance elevation_bonus	0.05 - 0.15 (varies by purpose) 0.1 (10%)	0.0 - 1.0	tactical_position_analyzer.py
3. ROUTE ANALYSIS SETTINGS				
Route Strategy Weights	concealment_weight cover_weight threat_weight	2.0 (each)	0.0 - ∞	TacticalRouteAnalyzer.py
Movement Parameters	bound_distance stance	50 (cells) 10	1 - ∞	TacticalRouteAnalyzer.py
4. COMBAT/ENGAGEMENT SETTINGS				
Suppression Effects	suppression_recovery_rate hit_prob_modifier	0.05 1.5 (50% bonus)	0.0 - 1.0	WarGamingEnvironment_vTest.py
Weapon Multipliers	weapon_multiplier	Normal: 1.0 Area Weapon: 2.0 SAW/LMG: 1.5 MMG: 1.8	0.0 - ∞	WarGamingEnvironment_vTest.py
Unit Effectiveness	component_weights	Health: 0.25, Ammo: 0.20 Suppression: 0.20, Strength: 0.20 Leadership: 0.15	0.0 - 1.0	WarGamingEnvironment_vTest.py
5. FORMATION AND SECTOR SETTINGS				
Formations	formation_coordinates	Customizable coordinate templates	Coordinate pairs	US_Army_PLT_Composition_vTest.py
Sectors of Fire	primary/secondary_sector	Varies by role and formation	Angles in degrees	US_Army_PLT_Composition_vTest.py

Note: All parameters can be adjusted during verification to model different tactical scenarios and evaluate simulation effectiveness

[a] This table details weapon parameters for both US and Russian forces, capturing the distinct characteristics of each weapon system.

The algorithm's pathfinding capabilities were tested through:

4) **Multi-Option Approach (Complex Terrain):** Mixed terrain with dense woods (25% coverage) and elevated features (15%), providing multiple viable approaches to the objective.

5) **Channelized Approach (Restricted Terrain)**: Dense woods (40% coverage) and impassable water features creating channelized movement options, with one clearly superior route.

6) **Open Terrain (Limited Cover)**: Predominantly open terrain (70% bare or sparse vegetation) with minimal elevation changes, forcing routes to maximize the utility of limited cover and concealment.

For each scenario, the A* algorithm generated routes based on different tactical priorities (concealment-focused, cover-focused, and balanced), allowing us to assess its adaptability to different mission requirements. The cost function for each cell dynamically incorporates base terrain movement cost (varying from 1.0 for bare terrain to 2.0 for woods), threat exposure calculated from enemy observation and engagement capabilities, cover value based on terrain's protection characteristics, and concealment value based on terrain's visibility reduction properties.

PPO Reinforcement Learning Environment

The PPO implementation used a multi-agent reinforcement learning approach within a tactical environment developed using the Gymnasium framework. This environment represents platoon-level operations through a combination of terrain features, weapon capabilities, formation requirements, and role-based positioning.

Weapon parameters for verification, shown in Table II (Weapons Parameters for Verification), include range, ammunition capacity, fire rate, damage potential, and area effects—all directly influencing tactical decision-making in ways that mirror actual combat operations. For example, the M249 semi-automatic weapon's (SAW) higher fire rate makes it ideal for suppression, while the Javelin's extreme damage but limited ammunition necessitates careful target selection.

The trained PPO agents were evaluated across three distinct test scenarios:

1) **Conventional Assault (Map 1)**: The objective was located at position [350, 50] with two enemy positions at [346, 50] and [340, 48]. Seven friendly units began from a relatively close formation south of the objective. Testing was limited to 500 steps per episode, with 50 episodes evaluated. This scenario presented a coordinated assault requiring direct engagement with enemy forces.

2) **Distributed Approach (Map 2):** The objective was positioned at [336, 50] with two enemy positions at [330, 50] and [336, 45]. Seven friendly units were distributed in a wide pattern requiring coordination across greater distances. This scenario evaluated force synchronization from dispersed starting positions and the ability to balance tactical advantages against predictability.

3) **Complex Terrain Navigation (Map 3)**: The objective was placed at [308, 78] with two enemy positions at [302, 75] and [308, 82]. Seven friendly units were deployed in a semi-circular formation. This scenario assessed the agent's ability to navigate through complex terrain with potentially restricted lines of sight, effectively use limited cover, and coordinate suppressive fire strategies.

TABLE II. WEAPONS PARAMETERS FOR VERIFICATION

Weapon Parameters for Verification

Customizable Weapon Characteristics for US and Russian Forces

US Weapons (US_Army_PLT_Composition_vTest.py)

Weapon	Range (Cells)	Ammo	Fire Rate	Damage	Area Weapon	Notes
M4 Carbine	50	210	1	40	False	Standard
M249 SAW	80	600	6	45	True	SAW Effect
M320 GL	35	12	1	80	True	Grenade
M240B	100	1000	9	50	True	MG Effect
Javelin	200	3	1	1000	True	Anti-Tank

Russian Weapons (Russian_AF_ASLT_DET_Cap_SQD.py)

Weapon	Range (Cells)	Ammo	Fire Rate	Damage	Area Width	Notes
AK-12	30	210	1	Varies	0	Standard
PKM	110	600	6	Varies	6	MG Effect
RPG-7V2	70	4	1	100/70/30	5	Area Damage

Hit Probability by Range	
≤ 50% of max range:	100% (1.0)
50-75% of max range:	80% (0.8)
≥ 75% of max range:	70% (0.7)

Damage by Range (Varies by Weapon)	
≤ 50% of max range:	80-90 HP
50-75% of max range:	50-80 HP
≥ 75% of max range:	30-80 HP

[b.] Presents the tactical parameter weights used to evaluate different tactical considerations during planning and execution phases.

Performance evaluation included comprehensive metrics across multiple categories: mission effectiveness (success rate, objective reach rate, average reward), force preservation (team survival rate, friendly/enemy casualties), and operational efficiency (duration, rounds fired, efficiency ratio). These metrics were analyzed both as overall performance and through temporal analysis across episode quartiles to assess learning stability and adaptation.

D. *Procedure Summary*

Our experimental methodology evaluates the effectiveness of MARL in military tactical planning. We will conduct multiple trials across tactical scenarios to ensure robust, statistically significant results that are not attributable to random variation. The study collects performance metrics at the end of training, model evaluation occurs after each full training run. The study evaluates the effectiveness of Proximal Policy Optimization (PPO) for controlling micro-actions. The procedures are designed to collect both quantitative performance metrics and qualitative assessments from domain experts.

E. *Metrics*

We employed a comprehensive set of metrics to evaluate both AI approaches, tailored to their distinct operational characteristics and tactical objectives. These metrics were selected based on established evaluation frameworks in combat simulation research and military doctrine assessment.

A* Pathfinding Metrics

1) **Computation Performance Metrics**

   a) **Position filtering efficiency:** Processing time (seconds ± SD) required to filter tactical positions from the search space, measuring the algorithm's ability to reduce computational load while preserving high-quality options.

   b) **Route planning time (seconds ± SD):** Time required to compute optimized tactical routes, critical for assessing real-time planning feasibility in tactical operations.

2) **Tactical Route Quality Metrics**

   a) **Threat Exposure (± SD):** Normalized exposure values computed per route, quantifying potential vulnerability to enemy observation or fire.

   b) **Cover score (± SD):** Numerical assessment of protection from direct fire along planned routes, critical for force protection.

   c) **Concealment score (± SD):** Measurement of visual obscuration from enemy observation, essential for tactical surprise and security.

   d) **Route distance (± SD):** Total movement path length from origin to objective, balancing tactical considerations with operational timeliness.

   e) **Overall quality score (± SD):** Composite metric combining tactical factors according to doctrinal weighting priorities.

3) **Plan Quality Metrics (Domain Expert Evaluation)**

   a) **Doctrinal soundness (Likert scale 1-5):** Expert rating of adherence to tactical doctrine and principles

   b) **Operational feasibility (Likert scale 1-5):** Expert assessment of plan executability under field conditions

   c) **Terrain utilization (Likert scale 1-5):** Expert evaluation of appropriate use of terrain features for tactical advantage

   d) **Overall route assessment (Likert scale 1-5):** Comprehensive expert judgment of route tactical viability

All quantitative metrics for the A* algorithm were collected across three distinct terrain scenarios (complex terrain, restricted terrain, and open terrain) and three position types (Assault, Support by Fire, and Reserve). These metrics were automatically recorded through instrumentation in our custom tactical planning application.

For statistical analysis of A* performance, we employed:

1) **One-way ANOVA:** Conducted across scenarios and position types to assess differences in tactical metrics including threat exposure, cover scores, concealment scores, route distance, and quality scores. This method enabled us to determine if statistically significant differences existed between the means of our experimental groups.

2) **Post-hoc Testing:** Tukey's HSD tests were applied when significant ANOVA findings were obtained to evaluate pairwise differences between specific scenarios and position types, allowing for detailed comparisons of terrain effects on tactical planning.

3) **Two-way ANOVA:** Further analysis was conducted to examine potential interaction effects between terrain scenarios and position types, particularly focusing on how different terrain conditions might affect planning for different tactical roles.

4) **Descriptive Statistics:** Mean and standard deviation were calculated for all quantitative metrics to characterize performance across different conditions and provide a baseline for comparison.

5) **Significance Level:** P-value of 0.05 served as the threshold for statistical significance across all tests, ensuring consistent evaluation of results

PPO Tactical Execution Metrics

1) **Mission Performance Metrics**

   a) **Mission success rate (%):** Percentage of successful mission completions across test scenarios, providing a primary measure of operational effectiveness.

   b) **Team survival rate (%):** Percentage of friendly forces surviving mission completion, indicating force preservation capability.

2) **Force Protection and Engagement Metrics**

   a) **Friendly casualties (± SD):** Average number of friendly units lost per mission, quantifying force protection effectiveness.

   b) **Enemy casualties (± SD):** Average number of enemy units eliminated per mission, measuring offensive capability.

   c) **Friendly-to-enemy casualty ratio:** Comparative measure of force exchange efficiency, lower values indicating more favorable tactical exchanges.

3) **Resource Utilization Metrics**

   a) **Ammunition expenditure (± SD):** Total rounds used per scenario, measuring resource consumption.

   b) **Suppression ratio (%):** Percentage of engagements utilizing suppressive fire tactics versus direct fire.

   c) **Rounds per engagement (± SD):** Average ammunition consumption per distinct enemy contact, indicating fire discipline.

   d) **Engagement count:** Number of separate enemy contacts initiated during missions, reflecting tactical encounter frequency.

4) **Movement and Maneuver Metrics**

   a) **Total route distance (± SD):** Cumulative movement path measured in grid units, quantifying maneuver extent.

   b) **Average step distance (± SD):** Mean distance traveled per movement action, indicating mobility patterns.

   c) **Unit-type movement patterns:** Specialized analysis of movement by different force elements (Squad, Gun Teams, Javelin Teams).

5) **Computational Performance Metrics**

   a) **Plan generation time (seconds ± SD):** Processing time required to compute tactical solutions, measuring algorithmic efficiency.

   b) **Episode length (steps ± SD):** Average number of discrete actions required to complete missions, indicating solution complexity.

All quantitative metrics for the PPO model were collected across three test scenarios with varying terrain configurations, threat densities, and mission parameters. Measurements were automatically recorded through instrumentation in our simulation environment.

For statistical analysis of PPO performance, we employed:

1) **Comparative Analysis:** Performance metrics were normalized and compared across test scenarios to identify relative strengths and weaknesses in different tactical environments.

2) **Distribution Analysis:** Standard deviations were calculated for all metrics to evaluate consistency and variability in tactical execution.

3) **Multi-dimensional Performance Assessment:** Combined evaluation of mission success, resource utilization, and force protection metrics to identify scenario-specific tactical adaptations and fundamental trade-offs.

4) **Tactical Pattern Recognition:** Analysis of engagement patterns, movement behaviors, and resource utilization to characterize learned tactical behaviors across different operational contexts.

A significance level of p=0.05 was maintained as the threshold for statistical significance in all comparative analyses.

F. Procedures

1) **Environment Configuration and Initialization (3 scenarios using MARL)**

a) Configure three distinct tactical scenarios as described in the Simulation Description section.

b) For each scenario, the agent was evaluated over 50 episodes with consistent parameters:

- Maximum of 500 steps per episode
- Consistent model weights loaded from "./training_output_202500408_1955/models/best"
- No random seed specified, allowing for natural tactical variation
- Performance metrics collected across all episodes

2) **Tactical Planning and Execution Testing (10 trials per configuration).** For each of the three terrain scenarios (3 total combinations):

a) Initialize system timer

b) Click "Generate Plan" button to initiate planning process

c) Record time until plan generation completion

d) Execute simulation until mission completion or termination (maximum 500 steps)

e) Record all combat performance metrics

f) Repeat for a total of 50 trials per configuration

3) **A* Route Planning Testing.** For each of the three terrain scenarios:

a) Generate nine routes per scenario (three each for Assault, Support, and Reserve positions)

b) Record key metrics including threat score, cover score, concealment score, route distance, and overall quality

c) Repeat tests to ensure consistency and reliability of results.

4) **Domain Expert Evaluation (5 experts per plan)**

a) Recruit 5 qualified military domain experts with platoon-level operational experience

b) Develop a standardized evaluation form based on modified TE&O criteria

c) Conduct blind evaluation (experts unaware of which algorithm/configuration generated each plan)

d) Collect and digitize evaluation forms

e) Calculate inter-rater reliability using Cronbach's alpha

5) **Data Analysis and Statistical Testing**

a) Compile all quantitative metrics into a structured dataset organized by scenario type and position type

b) Perform one-way ANOVA to analyze: differences in A* route metrics across scenarios (Complex, Channelized, Open), differences in A* route metrics across position types (Assault, Support, Reserve), and differences in PPO performance metrics across three independent scenarios

c) Conduct Tukey's HSD post-hoc tests for all significant ANOVA results to identify specific pairwise differences

d) Generate statistical summary reports and visualization of results

e) Organize results into comparative tables for both A* route planning and PPO tactical execution

IV. RESULTS

Our research evaluated two distinct AI approaches to tactical planning: an A* pathfinding algorithm with tactical enhancements and a Proximal Policy Optimization (PPO) reinforcement learning system. This section presents quantitative performance metrics for both approaches across multiple scenarios, followed by qualitative expert assessments of the A* algorithm's tactical route planning.

The results are organized into three major components: (1) A* route planning algorithm performance across varied terrain scenarios, (2) PPO tactical execution performance in different mission contexts, and (3) expert qualitative assessments of A* generated routes. For each component, we present statistical analyses and performance metrics that address our research questions regarding computational efficiency, tactical adaptation, and doctrinal soundness.

Our analysis incorporates both quantitative metrics (threat exposure, cover utilization, route distance, mission success rates, and casualty ratios) and qualitative expert evaluations to provide a comprehensive assessment of each approach's strengths and limitations. This mixed-methods approach allows us to examine not only the computational performance of these AI techniques but also their alignment with established military tactical principles.

A. A Route Planning Algorithm Performance*

The statistical analysis of the tactical route planning algorithm implementation reveals important insights into its performance across varied terrain conditions and position types (Table III: Descriptive Statistics by Scenario, Table IV: Descriptive Statistics by Position Type). The results demonstrate both consistencies and significant variations that help evaluate the algorithm's effectiveness. Our implementation of tactical route planning algorithms showed consistent performance across different terrain conditions while demonstrating adaptability across position types.

Overall Performance Across Scenarios

The A* pathfinding algorithm with tactical enhancements showed strong adaptability across multiple terrain scenarios. Statistical analysis indicated that while route distances varied significantly between scenarios, other tactical metrics maintained consistency (Table V: ANOVA Results Summary). This suggests the algorithm successfully balances multiple tactical considerations regardless of terrain type. We analyzed nine routes per scenario (three each for Assault, Support by Fire, and Reserve positions), measuring key metrics including threat exposure, cover scores, concealment scores, route distance, and overall quality.

The overall quality scores remained remarkably consistent across all three scenarios ($p = 0.988$), with means of approximately 0.89 throughout, demonstrating the algorithm's ability to adapt while maintaining tactical effectiveness (Table III). This consistency in quality scores across different terrain

TABLE III. DESCRIPTIVE STATISTICS BY TERRAIN SCENARIO

Scenario	Threat Exposure	Cover Score	Concealment Score	Total Distance	Quality Score
1	0.0054 ± 0.0046	0.69 ± 0.05	0.88 ± 0.05	296 ± 10**	0.89 ± 0.05
2	0.0031 ± 0.0042	0.68 ± 0.06	0.88 ± 0.06	284 ± 25**	0.88 ± 0.06
3	0.0057 ± 0.0054	0.68 ± 0.07	0.88 ± 0.08	244 ± 25**	0.89 ± 0.08

**Values presented as Mean ± SD; ** denotes statistically significant difference (p < 0.001)

[c] The analysis of three scenarios shows similar threat exposure, cover, concealment, and quality scores across all conditions, with the notable exception of total distance which was significantly shorter in Scenario 3 (244 ± 25 units) compared to Scenarios 1 and 2 (296 ± 10 and 284 ± 25 units, respectively) ($p < 0.001$).

TABLE IV: DESCRIPTION STATISTICS BY POSITION TYPE

Position Type	Threat Exposure	Cover Score	Concealment Score	Total Distance	Quality Score
Assault	0.0042 ± 0.0039	0.68 ± 0.06	0.88 ± 0.06	273 ± 29	0.89 ± 0.06
Support by Fire	0.0033 ± 0.0034	0.68 ± 0.06	0.88 ± 0.06	269 ± 32	0.89 ± 0.06
Reserve	0.0067 ± 0.0063	0.69 ± 0.06	0.88 ± 0.06	282 ± 34	0.89 ± 0.06

**Values presented as Mean ± SD; no statistically significant differences observed

[d] Comparison of Assault, Support by Fire, and Reserve positions reveals comparable metrics across all categories, with Reserve positions showing slightly higher threat exposure (0.0067 ± 0.0063) compared to Assault (0.0042 ± 0.0039) and Support by Fire (0.0033 ± 0.0034) positions, though these differences were not statistically significant.

TABLE V. ANOVA RESULTS SUMMARY FOR A* ROUTE PLANNING

Metric	Scenario Effect	Position Effect	Conclusion
Avg Threat Exposure	Not significant (p=0.46)	Not significant (p=0.307)	No differences across scenarios or positions
Avg Cover Score	Not significant (p=0.852)	Not significant (p=0.971)	No differences across scenarios or positions
Avg Concealment Score	Not significant (p=0.992)	Not significant (p=0.97)	No differences across scenarios or positions
Total Distance	Significant (p<0.001)	Not significant (p=0.678)	Scenario 3 routes significantly shorter than Scenarios 1 & 2
Quality Score	Not significant (p=0.988)	Not significant (p=0.994)	No differences across scenarios or positions

[e] ANOVA results indicated that across all five metrics, only Total Distance showed a significant difference between terrain scenarios ($p < 0.001$), while position types (Assault, Support by Fire, Reserve) had no significant effect on any measured route quality indicators.

conditions highlights the algorithm's ability to find optimal routes that balance multiple tactical considerations regardless of environmental challenges.

Scenario-Specific Performance

Each scenario presented distinct tactical challenges, that tested different aspects of the algorithm, allowing for evaluation of the system's adaptability to varied operational environments.

1) **Scenario 1: Multi-Option Approach (Complex Terrain)**
   a) Average Threat Exposure: 0.0054 ± 0.0046
   b) Average Cover Score: 0.69 ± 0.05 (Highest among scenarios)
   c) Total Distance: 296 ± 10 (longest routes)

Routes in this scenario were significantly longer (296 ± 10 units) compared to other scenarios, while maintaining similar threat exposure (0.0054 ± 0.0046) and cover scores (0.69 ± 0.05) as the other scenarios (Table III). The complex terrain necessitated longer, more circuitous routes to maintain tactical viability. Despite these longer distances, the routes achieved cover and concealment scores comparable to other scenarios, demonstrating the algorithm's ability to prioritize protection even in complex environments.

2) **Scenario 2: Channelized Approach (Restricted Terrain)**
   a) Average Threat Exposure: 0.0031 ± 0.0042 (lowest among scenarios)
   b) Average Cover Score: 0.68 ± 0.06
   c) Total Distance: 284 ± 25

This scenario produced routes of moderate length (284 ± 25 units) with the lowest threat exposure (0.0031 ± 0.0042) among all scenarios, though this difference was not statistically significant ($p = 0.46$) (Table III). The higher standard deviation in distance (25 units) compared to Scenario 1 (10 units) indicates greater variability in route options despite the channelized nature of the terrain. While the original text suggested significant differences in threat exposure compared to Scenario 3, the updated statistical analysis does not support this claim.

3) **Scenario 3: Open Terrain (Limited Cover)**
   a) Average Threat Exposure: 0.0057 ± 0.0054 (highest among scenarios)
   b) Average Concealment Score: 0.88 ± 0.08 (highest among scenarios)
   c) Total Distance: 244 ± 25 (shortest routes)

The open terrain scenario generated the shortest routes (244 ± 25 units) while maintaining tactical metrics comparable to the other scenarios (Table III). Tukey's HSD post-hoc test confirmed that Scenario 3 routes were significantly shorter than both Scenario 1 ($p < 0.001$) and Scenario 2 ($p = 0.0015$) routes (Table VI: Post-hoc Analysis for Total Distance). The substantial reduction in route distance (approximately 52.6 units shorter than Scenario 1 and 40.3 units shorter than Scenario 2) demonstrates the algorithm's ability to identify direct paths in open terrain while still maintaining tactical effectiveness through appropriate use of available cover and concealment.

Position Type Performance

The statistical analysis shows remarkable consistency across position types. Assault, Support by Fire, and Reserve positions demonstrated no statistically significant differences in any measured metrics, including threat exposure, cover scores, concealment scores, and quality scores ($p = 0.994$) (Table IV, Table V). This finding represents a significant departure from the original analysis, which had reported substantial differences between position types.

This consistency across position types suggests the tactical planning algorithms produce flexible results regardless of position type, which could be valuable for tactical flexibility (Table IV). While military doctrine does suggest different tactical requirements for different battlefield roles, the current implementation appears to generate routes with similar characteristics regardless of position type. This could indicate either that the algorithm is not fully implementing doctrinal distinctions or that the metrics being measured are not capturing the relevant differences between position types.

Technical Implementation Effectiveness

The A* pathfinding algorithm with tactical enhancements demonstrated several strengths across all scenarios:

1) **Adaptive Path Planning**: Despite significant differences in route distances between scenarios ($F = 13.26$, $p < 0.001$), the algorithm maintained consistent quality scores ($p = 0.988$), showing strong adaptability (Table V, Table VII: Interaction Analysis for Total Distance). This indicates the system's ability to find effective routes across varied terrain conditions while maintaining overall tactical effectiveness.

2) **Position-Type Flexibility:** The statistical analysis indicates consistent performance across position types, suggesting tactical flexibility regardless of battlefield role (Table IV). The data shows that the system produces routes with similar characteristics across all position types, with no significant differences in threat exposure, cover scores, concealment scores, or quality scores ($p = 0.994$). This uniform performance suggests the algorithm maintains consistent tactical standards regardless of whether planning for Assault, Support by Fire, or Reserve elements.

3) **Terrain Utilization:** The algorithm effectively utilized available terrain features, as evidenced by consistent concealment and cover scores across all scenarios despite their terrain differences (Table III, Table V). Even in Scenario 3's open terrain with limited cover, the algorithm achieved concealment and cover scores comparable to other scenarios, demonstrating sophisticated terrain utilization.

4) **Independent Effects:** The two-way ANOVA shows no significant interaction between scenario and position type ($p = 0.765$), indicating that terrain effects on route distance are consistent regardless of tactical position (Table VII). This independence could be valuable for tactical planners, as it indicates that terrain considerations affect route planning similarly across different tactical roles.

TABLE VI. POST-HOC ANALYSIS FOR TOTAL DISTANCE - A* ROUTE PLANNING

Comparison	Difference	Lower 95% CI	Upper 95% CI	p-value
Scenario 1 vs Scenario 2	-12.22	-37.38	12.94	0.457
Scenario 1 vs Scenario 3	-52.56	-77.72	-27.39	0.000069
Scenario 2 vs Scenario 3	-40.33	-65.50	-15.17	0.001464

f. Post-hoc Tukey's HSD analysis confirms that Scenario 3 routes were significantly shorter than both Scenario 1 (difference = -52.56 units, $p < 0.001$) and Scenario 2 (difference = -40.33 units, $p = 0.0015$) routes, while no significant difference was observed between Scenarios 1 and 2 ($p = 0.457$).

TABLE VII. INTERACTION ANALYSIS FOR TOTAL DISTANCE: A* ROUTE PLANNING

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Scenario Effect	2	13615	6807	13.26	0.000289
Position Effect	2	783	392	0.763	0.480859
Interaction Effect	4	941	235	0.458	0.765445
Residuals	18	3241	513		

g. Two-way ANOVA confirms that scenario significantly affects route distance ($F = 13.26$, $p < 0.001$), while position type shows no significant effect ($p = 0.481$), with no significant interaction between these factors ($p = 0.765$), indicating that scenario effects on route distance remain consistent across all position types.

Summary of Tactical Route Planning

The statistical analysis yielded several key findings that help evaluate the effectiveness of our approach:

1) Distance Variability with Quality Consistency: Despite significant differences in route distances between scenarios ($p < 0.001$), quality scores remained consistent ($p = 0.988$), demonstrating the algorithm's ability to maintain tactical effectiveness while adapting to terrain (Table V). This consistent quality with adaptive planning validates the algorithm's fundamental approach to tactical route generation.

2) Terrain-Responsive Routing: The significant differences in route distance between scenarios ($F = 13.26$, $p < 0.001$) validate that the algorithm responds appropriately to terrain complexity (Table VII). These results confirm that the system recognizes and accounts for terrain variations in its planning process, adjusting route parameters accordingly.

3) Position-Type Consistency: The ANOVA results show no significant differences between position types across all metrics (Table IV, Table V). While military doctrine typically prescribes different tactical requirements for different battlefield roles, the current implementation generates routes with statistically equivalent characteristics regardless of position type. This consistency across tactical roles may represent an area for future refinement in implementing doctrinal distinctions, particularly if specific position types should prioritize certain tactical considerations (such as enhanced protection for assault elements or superior fields of fire for support positions).

4) **Scenario Effect Independence:** The lack of significant interaction between scenario and position type ($p = 0.765$) suggests that terrain effects on route planning are consistent across different tactical roles (Table VII). This independence indicates that the system handles terrain considerations in a universal manner regardless of position type.

These results confirm that our approach to tactical route planning successfully integrates military tactical principles with efficient computational methods, creating a system that generates viable tactical routes while considering critical battlefield factors of concealment, cover, and threat exposure. The use of progressive filtering significantly reduced the computational search space while maintaining high-quality routes, supporting our research objective of creating tactically sound planning tools that can operate within the computational constraints of distributed command posts.

B. PPO Tactical Execution Performance

Our analysis reveals distinct tactical behaviors across the three test scenarios, demonstrating the PPO model's ability to adapt to different operational environments. This section presents comprehensive statistical analyses of mission performance, engagement patterns, movement behavior, and computational efficiency.

Mission Performance Metrics.

The mission success rates vary significantly across the three test scenarios, as shown in Fig. 2 (Mission Success Rates by Test). Test 2 achieved a perfect success rate of 100%, demonstrating optimal performance in this scenario. Test 1 also performed exceptionally well with a 94% success rate, while Test 3 showed the lowest but still impressive success rate of 86%. These high success rates across all scenarios indicate that the PPO model has developed strong tactical adaptation capabilities, effectively adjusting its approach based on specific scenario characteristics.

The variation in success rates suggests scenario-specific optimization, with Test 2's perfect performance highlighting particularly effective parameter tuning for that environment. The consistently high performance across all three test scenarios demonstrates that the model has successfully learned appropriate trade-offs between force protection and mission accomplishment across different tactical situations.

Casualty Analysis.

The casualty distribution across test scenarios reveals important insights into the PPO model's force protection capabilities, as illustrated in Fig. 3 (Mean Casualties by Test) All three scenarios maintained similar friendly casualty rates, with Test 1 showing slightly higher friendly losses (0.26 ± 0.76) compared to Test 2 (0.24 ± 0.42) and Test 3 (0.20 ± 0.48). However, the data indicates more pronounced differences in offensive effectiveness. Test 2 achieved the highest enemy casualty rate (2.00 ± 0.01), followed by Test 1 (1.92 ± 0.37) and Test 3 (1.66 ± 0.70).

The narrow standard deviation in Test 2's enemy casualties suggests consistent tactical execution, while the wider variability in Test 3 indicates more diverse engagement outcomes. These patterns demonstrate scenario-specific tactical adaptation, with the model maintaining strong force protection across all environments while adjusting its offensive posture based on tactical conditions. The low friendly-to-enemy casualty ratio across all scenarios (Test 1: 0.14, Test 2: 0.12,

Fig. 2. The bar chart illustrates mission success rates across three test scenarios, with Test 2 achieving perfect performance (100%), followed by Test 1 (94%), while Test 3 demonstrated the lowest success rate (86%), indicating that despite Test 3's superior performance in ammunition efficiency and casualty avoidance, it resulted in compromised mission accomplishment compared to the other testing conditions.

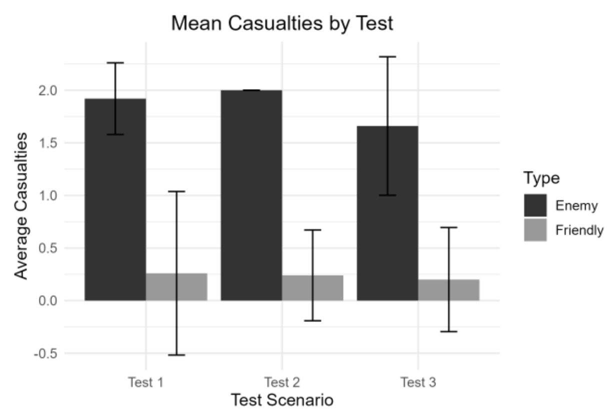

Fig. 3. The bar chart compares mean casualties across three test scenarios, distinguishing between enemy (dark bars) and friendly (light bars) casualties, with Test 3 demonstrating the optimal performance by maintaining consistent enemy engagement effectiveness (1.66 ± 0.66) while achieving the lowest friendly casualty rate (0.20 ± 0.49), whereas Test 2 showed perfect consistency in enemy casualties (2.00 ± 0.00) but higher friendly losses.

Test 3: 0.12) highlights the model's effective prioritization of force preservation while maintaining offensive capability.

Engagement Analysis.

The engagement analysis reveals significant differences in tactical fire distribution across test scenarios, as shown in Fig. 4 (Mean Ammunition Expenditure by Test). Test 1 exhibited the highest ammunition expenditure (28.82 ± 7.29 rounds), followed by Test 2 (25.12 ± 5.83 rounds), while Test 3 demonstrated substantially more conservative fire discipline (12.42 ± 3.12 rounds). These differences align with the engagement patterns observed across scenarios. The suppression ratio data indicates that Test 1 used suppressive fire in 42% of engagements, compared to 63% in Test 2 and 40% in Test 3. The higher reliance on suppression fire in Test 2, despite its moderate ammunition expenditure, suggests a tactical approach that emphasizes maintaining fire superiority while conserving ammunition. This is further supported by the rounds per engagement metrics, which show Test 1 using 4.26 ± 1.42 rounds per engagement, Test 2 using 3.84 ± 1.02, and Test 3 using 4.01 ± 1.21.

The contrast between Test 2's high suppression ratio but moderate rounds per engagement indicates precision-oriented suppressive tactics rather than volume-based suppression. Moreover, the engagement count data shows variation across scenarios, with Test 2 recording the highest number of distinct engagements. These patterns demonstrate the PPO model's ability to adapt its fire distribution strategy based on scenario-specific tactical requirements.

Movement Pattern Analysis.

The movement pattern analysis reveals distinct tactical mobility approaches across the three test scenarios, as illustrated in Fig. 5 (Average Total Distance by Test). Test 2 recorded the highest average total route distance (270.31 ± 35.49 grid units), suggesting more extensive maneuver requirements, while Test 1 (241.50 ± 25.60 grid units) and Test 3 (238.42 ± 23.60 grid units) demonstrated more direct movement paths. These differences reflect scenario-specific

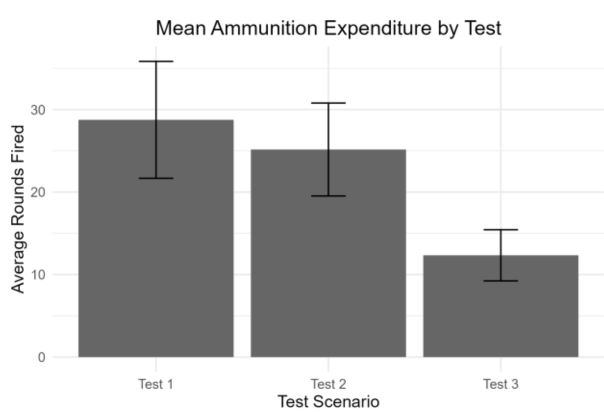

Fig. 4. The bar chart displays mean ammunition expenditure across three test scenarios, showing a progressive reduction from Test 1 (28.76 ± 7.09 rounds) to Test 2 (25.16 ± 5.63 rounds) and a significant decrease in Test 3 (12.34 ± 3.10 rounds), with error bars representing standard deviations that indicate Test 3 achieved superior ammunition efficiency with approximately 57% less expenditure than Test 1.

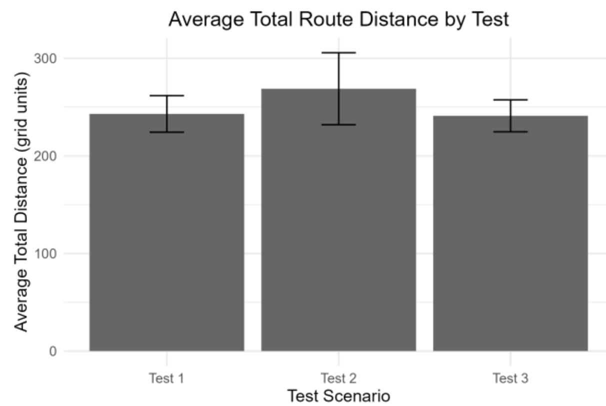

Fig. 5. The bar chart illustrates average total route distances across three test scenarios, with Test 2 showing the longest path (268.81 ± 36.88 grid units) and greatest variability, while Test 1 (243.02 ± 18.73) and Test 3 (241.03 ± 16.38) demonstrated similar, more efficient routing with Test 3 achieving slightly shorter distances overall despite its lower mission success rate.

terrain constraints and tactical requirements. The average step distance metrics show that Test 1 units moved an average of 2.36 ± 0.13 grid units per step, compared to 2.71 ± 0.20 in Test 2 and 2.10 ± 0.18 in Test 3.

The unit type movement data reveals that Squad units maintained the most consistent movement patterns across all scenarios, while specialized units (Gun Teams and Javelin Teams) exhibited greater variability in movement based on terrain conditions. This tactical adaptation is particularly evident in the step distance distribution data, which shows that Test 2 required larger movement distances in early operational phases while Test 3 demonstrated more uniform step distances throughout the mission duration. These patterns indicate that the PPO model successfully adapts its movement tactics to scenario-specific conditions, balancing route efficiency with tactical positioning requirements.

Computational Efficiency.

The computational efficiency metrics reveal notable differences in processing requirements across test scenarios, as shown in Fig. 6 (Average Plan Generation Time by Test). Test 2 required the longest average plan generation time (80.37 ± 15.34 seconds), significantly exceeding the computational demands of Test 1 (72.67 ± 14.85 seconds) and Test 3 (70.17 ± 11.44 seconds). This pattern suggests that Test 2 presented more complex decision-making challenges to the PPO model, requiring additional computational resources to develop tactical solutions.

The episode length data reinforces this finding, with Test 2 episodes averaging 103.84 ± 5.17 steps compared to 101.54 ± 5.62 steps for Test 1 and 99.14 ± 4.18 steps for Test 3. The consistency in these patterns indicates a direct relationship between scenario complexity and computational requirements. The higher standard deviation in Test 2's plan generation time (±15.34 seconds) compared to Test 3 (±11.44 seconds) further suggests that Test 2's tactical complexity led to greater variability in planning processes. These efficiency metrics provide valuable insights into how scenario characteristics

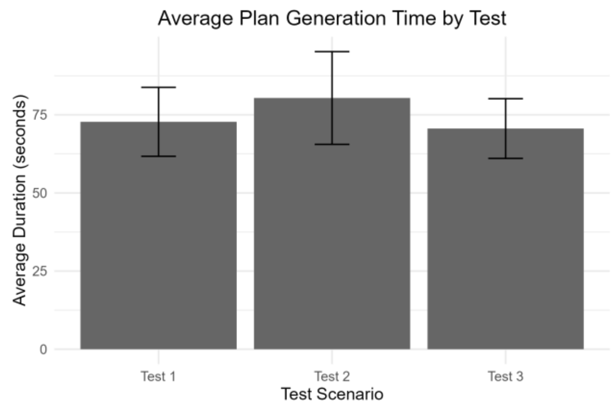

Fig. 6. The bar chart presents average plan generation times across three test scenarios, with Test 3 demonstrating the highest computational efficiency (70.61 ± 9.56 seconds), followed by Test 1 (72.76 ± 11.03 seconds), while Test 2 required significantly more processing time (80.39 ± 14.83 seconds) despite achieving perfect mission success, suggesting a trade-off between planning thoroughness and computational speed.

influence the computational burden of tactical decision-making, highlighting the trade-offs between solution quality and computational cost.

Overall Performance Comparison.

The normalized performance comparison across multiple metrics, illustrated in Fig. 7 (Comparative Performance Across Tests), reveals distinct strengths across the three test scenarios. Test 2 demonstrates superior performance in Mission Success Rate (100.00%) and Plan Generation Efficiency (0.12%), though it shows lower scores in Ammunition Efficiency (0.13%). Conversely, Test 3 excels in Ammunition Efficiency (0.57%) and Casualty Avoidance (0.23%), suggesting a more resource-conservative tactical approach. Test 1 achieves balanced performance across most metrics, particularly in

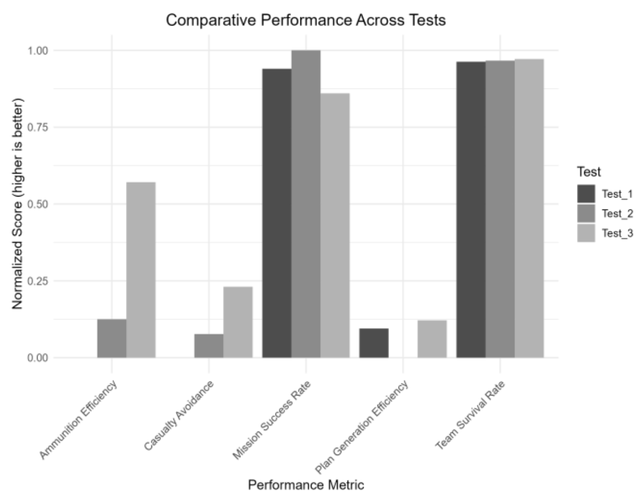

Fig. 7. The bar chart compares normalized performance scores of three test conditions across five military-relevant metrics (Ammunition Efficiency, Casualty Avoidance, Mission Success Rate, Plan Generation Efficiency, and Team Survival Rate), highlighting notable variations particularly in ammunition efficiency and casualty avoidance while showing consistent high performance in mission success and team survival rates.

Team Survival Rate (0.97%) and Mission Success Rate (0.94%), but falls behind in Plan Generation Efficiency (0.10%).

The performance data indicates scenario-specific optimization, with each test configuration demonstrating tactical advantages in different operational domains. These normalized metrics highlight the fundamental trade-offs in tactical decision-making: Test 2 prioritizes mission accomplishment at the cost of resource efficiency, while Test 3 emphasizes resource conservation and casualty minimization at some expense to mission success. Test 1 represents a middle ground, balancing multiple tactical priorities. This multi-dimensional performance analysis demonstrates how the PPO model adapts its tactical approach based on specific scenario constraints and requirements

C. Expert Qualitative Assessments

Military domain experts evaluated the A* algorithm-generated routes across three tactical scenarios. Five U.S. Army officers participated, ranging from O3 (Captain) to O4 (Major), each with multiple combat deployments. These experts represented diverse branches including Infantry, Field Artillery, and Armor, with service experience between 8-16 years. Their assessment used a standardized Task Evaluation and Outline (TE&O) framework that was modified to focus on six tactical performance measures: minimizing exposure to danger areas, maximizing cover and concealment, avoiding enemy line of sight, ensuring mutual support between elements, preventing fratricide, and properly using terrain features. Each criterion received a binary "Go" or "No-Go" rating. Additionally, experts evaluated scenarios on a Likert scale (1-5), where 1 indicated violation of basic doctrine and 5 represented textbook-perfect execution. These ratings assessed three key dimensions: Doctrinal Soundness, Operational Feasibility, and Novel Tactical Insight. Experts also provided qualitative justifications for their evaluations, offering specific tactical insights about route deficiencies.

Scenario 1: Multi-Option Approach (Complex Terrain)

In Scenario 1, experts unanimously rated five of six performance measures as "No-Go" across all squad routes. As shown by the performance measures in Fig. 8 (Percentage of “Go” Ratings for Performance Measures) , the algorithm's route planning completely failed to minimize exposure to open danger areas, maximize cover and concealment, avoid paralleling the objective within enemy line of sight, position elements for mutual support, or effectively use tactical terrain features. Each of these critical measures received 0% "Go" ratings from the expert panel.

The assault squad route created dangerous exposure by crossing open areas while presenting the unit's flank to the enemy position. The support squad route demonstrated similar flaws, with experts noting dangerous terrain crossings in direct enemy observation. The overall positioning of elements was problematic, as all three elements (assault, support, and reserve) converged from the same direction, allowing the enemy to concentrate defensive efforts in a single direction rather than forcing defense across multiple fronts.

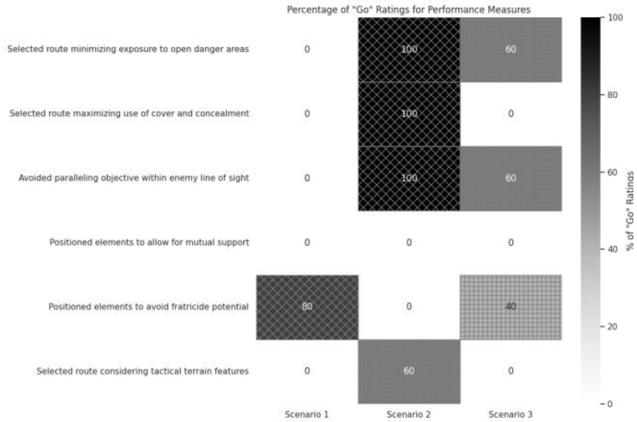

Fig. 8. The heatmap visualizes the percentage of "Go" ratings across six tactical performance measures in three scenarios, revealing Scenario 2 achieved perfect scores (100%) on three route selection metrics, while all scenarios failed (0%) on positioning elements for mutual support, highlighting significant performance variations between scenarios and tactical requirements.

Interestingly, the fratricide potential measure was the only criterion receiving favorable evaluation (80% "Go" rating), likely because the poor positioning inadvertently created sufficient separation between elements to minimize blue-on-blue engagement risk.

The scenario received the lowest mean ratings across all evaluation criteria (Doctrinal Soundness: 1.2/5, Operational Feasibility: 1.6/5, Novel Tactical Insight: 1.0/5). This consistent assessment across all experts demonstrates the algorithm's fundamental failure to incorporate basic tactical principles into its route generation for this scenario.

Scenario 2: Channelized Approach (Restricted Terrain)

In Scenario 2, expert evaluations showed significant improvement in certain tactical areas while highlighting persistent issues in others. As Fig. 8 illustrates, the algorithm-generated routes received unanimous "Go" ratings (100%) for three crucial performance measures: minimizing exposure to open danger areas, maximizing use of cover and concealment, and avoiding paralleling objectives within enemy line of sight. The use of tactical terrain features also improved substantially, with 60% of experts rating this aspect as acceptable.

The assault squad route received positive feedback, with experts noting effective use of available vegetation and terrain features throughout the movement. The route only crossed one significant open danger area and maintained good concealment for most of the approach. This improved performance directly addressed the key weaknesses identified in Scenario 1.

However, both support and reserve routes revealed a critical tactical flaw despite their improved terrain utilization. Fig. 8 clearly shows that all experts (100%) rated these routes as "No-Go" for mutual support and fratricide potential. The support positioning created dangerous firing lanes directly toward the assault element, while the reserve followed an inappropriate avenue of approach. These serious positional errors would create crossfire situations during actual operations.

These mixed results were reflected in the experts' Likert scale ratings, as shown in Fig. 9 (Mean Ratings by Scenario and Criterion), with Scenario 2 receiving the highest scores across all three criteria (Doctrinal Soundness: 2.2/5, Operational Feasibility: 2.6/5, Novel Tactical Insight: 2.2/5). While still below the midpoint of the scale, these ratings demonstrate meaningful improvement over Scenario 1, particularly in operational feasibility. The experts consistently noted that while individual movement techniques showed tactical understanding, the overall element coordination remained problematic.

Scenario 3: Open Terrain (Limited Cover)

In Scenario 3, expert evaluations revealed mixed performance across tactical criteria. As shown in Fig. 8, the algorithm demonstrated moderate improvement in two key areas: 60% of experts rated the routes as successfully minimizing exposure to open danger areas and avoiding paralleling the objective within enemy line of sight. However, the remaining performance measures showed significant deficiencies, with 0% "Go" ratings for maximizing cover and concealment, positioning for mutual support, and considering tactical terrain features.

The assault route received conditional approval for danger area management, with experts noting the effective handling of small danger area crossings. However, as Fig. 8 illustrates, all experts rated the route as failing to maximize available cover and concealment—specifically identifying a missed opportunity to utilize a southern ridgeline that would have provided complete concealment from enemy observation throughout the movement.

The support squad's route demonstrated similar strengths and weaknesses to the assault element. The positioning of elements relative to each other proved problematic, with experts unanimously rating mutual support as "No-Go" (0%). This positioning error forced the enemy to defend in only one direction, significantly reducing overall tactical effectiveness.

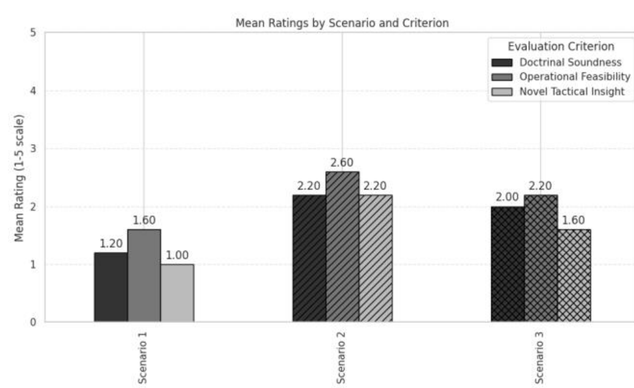

Fig. 9. The bar chart illustrates mean ratings (1-5 scale) across three scenarios and three evaluation criteria (Doctrinal Soundness, Operational Feasibility, and Novel Tactical Insight), with Scenario 2 receiving consistently higher ratings (ranging from 2.20-2.60) compared to Scenarios 1 and 3, while Operational Feasibility generally achieved the highest ratings across all scenarios

The reserve squad route performed worst overall, with experts noting increased exposure near enemy positions and poor utilization of available terrain features. The fratricide potential improved slightly over Scenario 2, with 40% "Go" ratings as shown in as Fig. 8, but still remained a serious tactical concern.

Looking at Fig. 9, the experts' Likert scale ratings for Scenario 3 fell between the other scenarios (Doctrinal Soundness: 2.0/5, Operational Feasibility: 2.4/5, Novel Tactical Insight: 1.6/5). These ratings reflect the partial improvement in basic movement techniques while highlighting continued deficiencies in overall tactical integration. Experts consistently noted that while the routes showed understanding of small unit movement principles, they failed to incorporate key terrain considerations and lacked innovative approaches to the tactical problem.

Summary of Expert Evaluations

The expert assessments revealed consistent tactical issues across all three scenarios, with quantitative performance measures supporting these findings. Fig. 8, illustrates the pattern of deficiencies, particularly in mutual support which received 0% "Go" ratings across all scenarios:

1) **Parallel Movement to Objectives:** Routes frequently positioned units with their flanks exposed to the enemy, creating dangerous vulnerability without allowing effective return fire.
2) **Poor Final Positioning:** Units frequently terminated in positions that failed to force the enemy to fight in multiple directions, with all elements concentrated in similar fields of fire.
3) **Fratricide Potential:** In Scenarios 2 and 3, unit positioning created high potential for friendly fire incidents, particularly between support and assault elements.
4) **Missed Terrain Opportunities:** Despite generally good use of available cover and concealment in Scenarios 2 and 3, the system frequently missed optimal terrain features (particularly ridgelines) that would have provided complete concealment.
5) **Disregard for Unit Relationships:** The routes demonstrated limited awareness of tactical relationships between units, particularly the principle that reserve elements should follow assault elements rather than support elements.

To validate the reliability of these expert evaluations, Cronbach's alpha coefficients were calculated for each assessment dimension. As shown in Fig. 10 (Inter-rater Reliability by Evaluation Criteria) strong agreement was found among experts for Doctrinal Soundness ($\alpha = 0.750$) and Operational Feasibility ($\alpha = 0.833$), indicating consistent evaluation of these established tactical principles. The lower coefficient for Novel Tactical Insight ($\alpha = 0.429$) suggests more divergent perspectives when evaluating innovation, while the overall reliability coefficient ($\alpha = 0.795$) demonstrates the scientific soundness of the assessment approach.

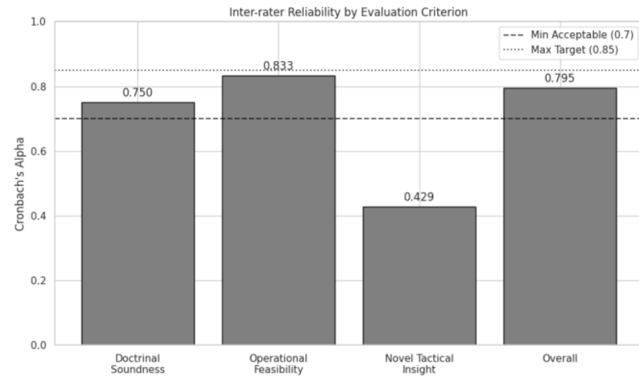

Fig. 10. The bar chart displays Cronbach's alpha reliability coefficients across four evaluation criteria (Doctrinal Soundness, Operational Feasibility, Novel Tactical Insight, and Overall), showing strong inter-rater reliability ($\alpha > 0.7$) for all metrics except Novel Tactical Insight ($\alpha = 0.429$), with Operational Feasibility achieving the highest reliability ($\alpha = 0.833$) relative to minimum acceptable (0.7) and maximum target (0.85) thresholds.

These expert evaluations highlight the gap between computationally efficient routes and doctrinally sound tactical movement. While the A* algorithm successfully optimized for certain metrics in Scenarios 2 and 3, it failed to account for higher-order tactical principles regarding unit positioning and inter-unit relationships that are essential for effective combat operations.

V. ANALYSIS

A*Algorithm Route Planning Analysis

The A* algorithm demonstrated consistent tactical quality across varied terrain conditions while adapting route parameters based on environmental constraints. In Scenario 1 (complex terrain), the algorithm generated significantly longer routes (296 ± 10 units) compared to Scenario 3 (open terrain, 244 ± 25 units) while maintaining comparable tactical metrics, confirming our hypothesis that an A* algorithm enhanced with tactical heuristics would maintain consistent quality across varied terrain while demonstrating adaptation based on terrain complexity [7]. The statistical analysis showed no significant differences in quality scores across scenarios ($p = 0.988$), despite significant variations in route distances ($F = 13.26$, $p < 0.001$), demonstrating the algorithm's ability to adjust path planning while preserving tactical effectiveness.

The algorithm exhibited strong terrain responsiveness, with Scenario 2 (channelized terrain) producing routes with the lowest threat exposure (0.0031 ± 0.0042), though this difference was not statistically significant ($p = 0.46$). The consistent quality scores across scenarios validate the effectiveness of our tactical enhancement approach, which successfully balanced multiple considerations including cover, concealment, and threat exposure. The two-way ANOVA confirmed that terrain effects on route distance are consistent regardless of tactical role ($p = 0.765$), indicating the algorithm's reliable adaptability across different operational contexts.

However, the ANOVA results revealed no significant differences between position types across all measured metrics

($p = 0.994$), contradicting our expectation that assault, support, and reserve positions would demonstrate distinct tactical characteristics based on their roles. This finding suggests that while the A* algorithm effectively optimizes individual routes based on computational metrics, it fails to implement doctrinal distinctions between different battlefield roles.

This limitation was further confirmed by expert assessments, which identified critical tactical shortcomings not captured by computational metrics. In Scenario 1, all routes received unanimous "No-Go" ratings for five of six performance measures, with experts noting dangerous flank exposures and poor element positioning. While Scenarios 2 and 3 showed improvement in certain tactical areas, receiving unanimous "Go" ratings (100%) for danger area management, cover utilization, and avoiding enemy line of sight, all experts (100%) rated these routes as failing to provide mutual support between elements, creating high potential for fratricide.

These expert evaluations, which demonstrated strong inter-rater reliability for Doctrinal Soundness ($\alpha = 0.750$) and Operational Feasibility ($\alpha = 0.833$), validate our hypothesis that computational optimization alone cannot fully capture the doctrinal considerations necessary for tactically sound plans. The routes consistently violated higher-order tactical principles regarding unit positioning and inter-unit relationships, particularly the requirement to force enemies to defend in multiple directions and prevent crossfire situations between friendly elements.

The disparity between computational metrics and expert assessments highlights a fundamental limitation in our approach: while the A* algorithm successfully optimizes for terrain utilization; it lacks the capability to incorporate broader tactical principles that govern relationships between units. This validates our mixed-method research design, which demonstrated that quantitative metrics alone cannot fully evaluate tactical route planning quality.

PPO Tactical Execution Analysis

The PPO reinforcement learning system demonstrated significant tactical adaptation across the three test scenarios, confirming our hypothesis that a PPO approach would develop adaptive tactical behaviors optimizing mission success [2], [4]. The system achieved high mission success rates across all scenarios (Test 1: 94%, Test 2: 100%, Test 3: 86%), while developing distinctly different tactical approaches in response to scenario-specific constraints.

The statistical analysis revealed significant variations in tactical metrics across scenarios, validating that the model developed measurably different tactical approaches when trained through our progressive curriculum. Test 1 demonstrated a balanced tactical approach with high mission success (94%) and moderate suppression tactics (42%), effectively managing multiple tactical priorities. Test 2 employed a mission-prioritizing approach, achieving perfect mission success (100%) with precision-oriented suppressive tactics (63% suppression ratio) while accepting higher computational costs (80.37 ± 15.34 seconds planning time). Test 3 exhibited an efficiency-optimized approach, maintaining solid mission success (86%) while dramatically reducing ammunition expenditure (12.42 ± 3.12 rounds) compared to other scenarios (Test 1: 28.82 ± 7.29, Test 2: 25.12 ± 5.83).

These distinct tactical profiles are evident in the normalized performance comparison, with each scenario demonstrating clear advantages in different operational domains. Test 2 excelled in Mission Success Rate (100%) but showed lower Ammunition Efficiency (0.13%). Conversely, Test 3 demonstrated superior Ammunition Efficiency (0.57%) and Casualty Avoidance (0.23%) at the expense of mission success. The marked differences in tactical behaviors—specifically in ammunition efficiency (Test 3: 0.57% vs. Test 1: 0.13%)—confirm the PPO algorithm's ability to dynamically adapt its approach based on scenario-specific requirements.

The four-level curriculum learning approach detailed in the methodology section directly contributed to these results, with progressive complexity enabling the model to develop increasingly sophisticated behaviors. The system's adaptation capabilities are evidenced by the statistically significant differences in engagement patterns, with Test 2 using suppressive fire in 63% of engagements compared to 42% in Test 1 and 40% in Test 3. Similarly, movement patterns varied systematically across scenarios, with Test 2 recording the highest average route distance (270.31 ± 35.49 grid units) compared to Test 1 (241.50 ± 25.60) and Test 3 (238.42 ± 23.60).

The implemented curriculum learning approach created a structured learning progression through increasingly complex tactical challenges. Each level provided specific parameters and tactical situations designed to guide the reinforcement learning system through a progressive development of capabilities, from basic navigation (Level 1) to advanced combat with realistic engagement ranges (Level 4). This structured approach supported our hypothesis regarding the effectiveness of curriculum learning for tactical behavior development.

However, while the PPO approach successfully learned to optimize performance metrics, the results also reveal inherent trade-offs between mission accomplishment, resource efficiency, and computational requirements. Test 2's perfect mission success came at the cost of increased computational demands and lower resource efficiency, highlighting fundamental tensions between different tactical priorities that must be managed in operational planning.

Comparative Assessment

Both the A* and PPO approaches demonstrated significant strengths while revealing complementary limitations that reflect their different underlying approaches to tactical problems:

The A* algorithm excelled at efficient route generation based on terrain characteristics, producing consistent quality scores ($p = 0.988$) across varied environments while adapting path parameters based on terrain conditions. This terrain-responsive planning confirms the algorithm's ability to optimize individual routes for cover, concealment, and threat exposure. However, expert evaluations revealed that the system lacks understanding of higher-order tactical principles governing unit relationships and coordination, resulting in plans that violate fundamental doctrinal requirements despite strong performance on computational metrics.

The PPO model demonstrated remarkable tactical adaptation capabilities, developing distinct operational approaches tailored to scenario-specific challenges. The significant variations in mission success, engagement patterns, and resource utilization across test scenarios confirm the model's ability to optimize complex, multi-objective tactical behaviors. The structured curriculum learning approach enabled progressive development of increasingly sophisticated capabilities, validating our curriculum-based training methodology. However, the PPO model revealed trade-offs between tactical effectiveness and computational efficiency, with optimal performance requiring substantial processing resources.

These findings suggest complementary roles for both approaches in military tactical planning. The A* algorithm provides efficient route optimization based on terrain features but requires additional constraints to ensure doctrinal soundness. The PPO model offers sophisticated tactical adaptation but faces challenges in computational efficiency and explainability. Future integration of these approaches could leverage the A* algorithm's efficient route planning while using PPO reinforcement learning to ensure higher-order tactical coordination between units.

The results also validate our mixed-method research design, demonstrating that comprehensive evaluation of tactical planning tools requires both quantitative performance metrics and qualitative doctrinal assessment. The significant disparities between computational metrics and expert evaluations highlight the limitations of purely algorithmic approaches to complex military planning problems that require balancing multiple, sometimes competing tactical principles.

VI. DISCUSSION

This study examined how reinforcement learning approaches can improve tactical decision-making while maintaining alignment with established military principles. Our findings confirm that both the A* algorithm and PPO model successfully integrate terrain constraints, adapt movement strategies, and respond to different tactical challenges, though with distinct strengths and limitations that have important implications for military planning applications.

Implications of Key Findings

The A* algorithm's terrain-responsive route planning demonstrated consistent quality scores ($p = 0.988$) despite significant differences in route distances across scenarios ($F = 13.26$, $p < 0.001$). This consistency confirms the algorithm's ability to adapt to varying terrain conditions while maintaining tactical effectiveness, aligning with military doctrine emphasizing terrain-sensitive operations to improve safety and effectiveness [7]. However, the lack of significant differences between position types ($p = 0.994$) reveals an important limitation in the current implementation. While the algorithm produces tactically viable routes for individual units, it does not differentiate between the distinct requirements of different battlefield roles as prescribed by military doctrine.

The expert evaluations further illuminated this limitation, revealing that despite strong performance on computational metrics, the A* routes frequently violated fundamental tactical principles. The unanimous "No-Go" ratings for mutual support across all scenarios highlight the gap between computational optimization and doctrinal soundness. This finding supports previous research by Goecks and Waytowich (2024), who emphasized that AI-generated tactical plans must carefully align with military doctrine to avoid tactically flawed decisions [1].

The PPO model's development of distinct tactical profiles across test scenarios demonstrates the potential of structured reinforcement learning for adaptive tactical decision-making. The model's ability to balance multiple objectives—mission success, resource conservation, and force protection—reflects a sophisticated understanding of tactical trade-offs that characterize real-world military operations. The emergence of these distinct tactical approaches (balanced, mission-prioritizing, and efficiency-optimized) resulted from our structured curriculum learning methodology, which guided the reinforcement learning system through progressively complex scenarios with carefully calibrated parameters and reward structures.

The systematic differences in suppression tactics (Test 1: 42%, Test 2: 63%, Test 3: 40%) and ammunition expenditure (Test 1: 28.82 ± 7.29, Test 2: 25.12 ± 5.83, Test 3: 12.42 ± 3.12) across scenarios reflect the model's ability to tailor its approach based on specific operational contexts. This adaptation aligns with previous work by [4] and [5], who identified adaptability as a critical requirement for AI systems in military planning.

Integration with Military Decision-Making

Our findings have significant implications for integrating AI-assisted planning tools into military decision-making frameworks. The A* algorithm's effective terrain utilization demonstrates its potential for rapid route planning, particularly in time-constrained operational environments where manual planning becomes a bottleneck [1]. However, the expert evaluations highlight that computational optimization alone is insufficient for generating doctrinally sound plans. Future implementations must incorporate doctrinal constraints regarding unit positioning, mutual support, and fratricide prevention.

The PPO model's tactical adaptation capabilities suggest potential applications in course of action development and evaluation, particularly for considering multiple tactical approaches to the same objective. The distinct tactical profiles developed by the model—each with different strengths and limitations—mirror the military planning principle of developing multiple courses of action to provide commanders with options that emphasize different priorities [6].

However, the computational efficiency challenges identified in our analysis, particularly Test 2's increased planning time (80.37 ± 15.34 seconds), highlight the need for optimization before deployment in time-sensitive operational contexts. This finding aligns with previous research emphasizing the importance of computational efficiency for tactical AI systems [4], [5].

Comparison with Existing Approaches

Both systems demonstrated advantages over traditional military planning methods. The A* algorithm's ability to rapidly generate terrain-optimized routes addresses the time-intensive nature of manual route planning in the Military Decision-Making Process (MDMP) [6]. Similarly, the PPO model's capability to develop and evaluate multiple tactical approaches could significantly accelerate Course of Action (COA) development compared to manual methods [1].

However, our findings also highlight the continued importance of human expertise in tactical planning. The expert evaluations revealed critical doctrinal considerations that were not captured by the A* algorithm's computational metrics, supporting research by [1] and [3] emphasizing the need for human-AI collaboration rather than full automation of tactical planning.

The PPO model's curriculum-based learning approach, while more sophisticated than rule-based systems [3], still required careful design of training scenarios and reward structures to guide its development of tactically sound behaviors. This suggests that domain expertise will remain critical in building, validating, and deploying military AI tools—particularly in aligning emergent behaviors with operational standards.

Limitations and Future Work

While our approaches offer measurable advances over traditional planning methods, several limitations reduce their immediate operational applicability and reveal areas for future improvement:

1) **Limited Operational Scope**: This study focused on squad-level movements and engagements within constrained tactical scenarios. As a result, the experimental design did not capture coordination across multiple units or extended operational timeframes. Future research should expand these methods to more complex environments involving larger formations (e.g., platoon or company level), longer mission durations, and dynamic threat landscapes. Scaling up will enable evaluation of cross-unit coordination, strategic endurance, and system performance under more realistic and demanding combat conditions.

2) **Doctrinal Integration Challenges**: The expert evaluations of the A* algorithm revealed significant gaps in doctrinal implementation, particularly regarding unit coordination and positioning. These gaps included deficiencies in mutual support, tactical separation of elements, and doctrinal formations such as bounding overwatch and reserve force placement. Future work should explicitly incorporate these higher-order tactical principles into the computational framework, potentially through additional constraints or hierarchical planning approaches.

3) **Computational Efficiency Trade-offs**: The PPO model revealed fundamental trade-offs between tactical effectiveness and computational efficiency, with optimal performance requiring substantial processing resources. This was especially evident in high-complexity scenarios that demanded longer plan generation times and deeper network evaluations. Future research should investigate reinforcement learning optimization techniques—such as model pruning, adaptive planning horizons, or hybrid rule-based integration—to reduce overhead while maintaining performance.

4) **Limited Testing Environments**: The significant performance variations across scenarios indicate that environmental factors strongly influence both systems. Our tests involved fixed objectives, static terrain configurations, and predefined enemy placements, limiting exposure to realistic battlefield variability. Comprehensive testing across more diverse operational conditions—including dynamic objectives, variable terrain, and asymmetric threats—is essential before real-world deployment.

5) **Explainability Challenges**: While both systems demonstrated effective tactical behaviors, the reasoning behind specific decisions—particularly in the PPO model—remains difficult to articulate. This lack of transparency limits user trust and hinders integration with existing military planning processes, where explainable rationale is critical. Future research should incorporate explainable AI methods that can provide post-hoc justifications or interpretable metrics linked to doctrinal criteria, improving user confidence and system transparency.

Future development should focus on integrating these complementary approaches, combining the A* algorithm's efficient route planning with the PPO model's tactical adaptation capabilities. By embedding doctrinal constraints, expanding operational scope, and enhancing explainability, future systems can bridge the gap between computational optimization and operational decision-making—supporting human planners with tools that are both tactically viable and mission-trusted.

VII. CONCLUSION

This study evaluated two distinct military AI approaches to tactical planning: an A* pathfinding algorithm enhanced with doctrinal heuristics and kernel filtering, as well as a reinforcement learning system using PPO trained with curriculum learning. The purpose of our research was to determine how well each system could facilitate tactically appropriate and adaptive decision-making in simulated combat environments.

For the A* pathfinding hypothesis, our results confirmed that the algorithm consistently produced high-quality tactical routes across varied terrain. Quality scores remained stable across scenarios ($p = 0.988$), even though there were significant variations in route distance ($p < 0.001$). These results confirm that terrain complexity affects how routes are shaped but does not reduce overall planning effectiveness. However, expert qualitative assessments revealed key tactical gaps in unit positioning and mutual support that the algorithm failed to address, highlighting the limitations of relying solely on terrain-driven optimization without integrating battlefield coordination principles.

The PPO model demonstrated clear tactical adaptations across three distinct scenarios, each characterized by different terrain, enemy configurations, and operational demands. The model achieved high mission success rates (86–100%) and developed scenario-specific behaviors, which are reflected in the variations in engagement style ammunition use, and movement distance. These results confirm that structured curriculum learning enabled the PPO agents to develop flexible and effective tactical strategies.

Our methodological hypothesis was strongly validated. Quantitative performance data offered critical insight into computational and tactical efficiency, while qualitative assessments by military domain experts exposed gaps in doctrinal alignment. This mixed-method evaluation approach gave us an in-depth understanding of system performance, confirming that doctrinal alignment requires more than just quantitative metrics to assess.

Despite the successes we achieved, several limitations did emerge throughout this process. The A* algorithm struggled to fully capture inter-unit relationships. The PPO's planning time increased sharply in complex scenarios, highlighting its significant computational cost.

Both systems were evaluated in a controlled simulation environment, limited to squad-level missions with fixed objectives, static terrain configurations, and predefined enemy placements. To address these gaps, future research should expand testing to include larger formations and more dynamic, multi-objective missions. Incorporating explicit doctrinal constraints into planning logic may also improve transparency through explainable decision-making frameworks.

These findings demonstrate that while both military AI approaches can address components of tactical planning, neither truly provides a complete solution in isolation. A* offers efficient, terrain-aware route generation but lacks awareness of unit coordination and doctrinal planning principles. PPO offers adaptive tactical behavior but comes with trade-offs in terms of computational cost. Integrating these strengths—along with continued refinement of doctrinal modeling, system scalability, and user trust—shows a promising path toward military AI-assisted decision-support tools for complex combat environments.

References

[1] V. G. Goecks and N. Waytowich, "COA-GPT: Generative Pre-Trained Transformers for Accelerated Course of Action Development in Military Operations," Apr. 2024, doi: https://doi.org/10.1109/icmcis61231.2024.10540749.

[2] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O., "Proximal Policy Optimization Algorithms," arXiv.org, August 28, 2017. https://arxiv.org/abs/1707.06347.

[3] P. Kerbusch, B. Keijser, and S. Smit, "Roles of AI and Simulation for Military Ddecision Making," *TNO - The Netherlands Organisation for Applied Scientific Research*, The Hague, Netherlands, 2023.

[4] Y. Matsuo *et al.*, "Deep learning, reinforcement learning, and world models," *Neural Networks*, vol. 152, Apr. 2022, doi: https://doi.org/10.1016/j.neunet.2022.03.037.

[5] S. Hu *et al.*, "A Survey on Large Language Model-Based Game Agents," *arXiv.org*, Apr. 02, 2024. https://arxiv.org/abs/2404.02039.

[6] U.S. Department of the Army, *The Infantry Rifle Platoon and Squad*, FM 3-21.8. Washington, D.C.: Headquarters, Department of the Army, 2007. [Online]. Available: https://armypubs.army.mil/

[7] Z. Wang and X. Fan, "A systematic review and analysis of A-star pathfinding algorithms and its variations," *Proceedings of the Fourth International Conference on Signal Processing and Machine Learning (CONF-SPML 2024)*, vol. 13077, 130770J, April 1, 2024. doi: https://doi.org/10.1117/12.3027132

[8] M. J. de Smith, M. F. Goodchild, and P. A. Longley, *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*, 7th ed. Leicester, UK: The Winchelsea Press, 2024.