

A framework of large language model commander agent for spatial reasoning in combat simulation

Received: 8 December 2025

Accepted: 4 March 2026

Published online: 13 March 2026

Cite this article as: Chen Y., Ping Y., Zhou S. *et al.* A framework of large language model commander agent for spatial reasoning in combat simulation. *Sci Rep* (2026). <https://doi.org/10.1038/s41598-026-43365-3>

Yi-bo Chen, Yang Ping, Shuhang Zhou & Caleb Jojo

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A Framework of Large Language Model Commander Agent for Spatial Reasoning in Combat Simulation

Yi-bo Chen^{1,*}, Yang Ping², Shuhang Zhou³ and Caleb Jojo⁴

^{1,2}PLA Academy of Military Science, Beijing 100091, China

³North University of China, Taiyuan 030051, China

⁴Northwestern University, Evanston, IL 60208, USA

*Correspondence: 924926706@qq.com

Abstract: Large language models (LLMs) demonstrate strong reasoning and planning capabilities in static textual contexts, yet they struggle significantly with dynamic decision-making tasks involving spatial elements, such as point selection in military simulations. These limitations arise from their reduced capacity to integrate real-time geographic data and adapt to spatial conditions, which can lead to crucial errors in positioning decisions. Such deficiencies may result in missed opportunities for tactical advantages, increased vulnerability, and diminished overall effectiveness in combat scenarios. To mitigate these issues, this paper presents the Geo-Commander framework, an innovative multi-task agent to combat simulations by integrate the ReAct reasoning mechanism and spatial encoding. The Geo-Choice module of this framework employs hexagonal grid encoding for preliminary location screening, enabling the agent to establish spatial constraints early in the decision-making process. The ReAct chain of this framework incorporates detailed geographic insights into the reasoning loop, yielding interpretable decisions for point selection. We validate the framework through experiments that reveal substantial performance improvements in both static point selections and real-time dynamic command tasks within a tank detachment combat simulation environment. Results indicate that Geo-Commander consistently surpasses control groups across various metrics, including selection quality, win rate, and overall combat effectiveness. These performance metrics highlight the framework's potential to meet the demands of dynamic combat environments, ultimately confirming the feasibility of integrating spatial reasoning within LLM frameworks and opening avenues for advancements in multi-agent geospatial intelligence systems and battlefield decision-making support.

Keywords: VLLM Agent, ReAct, Spatial Reasoning, Combat Simulation, Commander Agent

1 Introduction

In recent years, Large Language Models (LLMs) have shown exceptional reasoning and planning capabilities across various tasks, ranging from text understanding and tool use [1] to autonomous actions in complex environments [2, 3]. However, their strengths are primarily concentrated in static text and symbolic reasoning scenarios, revealing significant limitations when applied to dynamic spatial tasks [4]. In particular, point selection in decision-making tasks necessitates not only understanding the spatial

structure but also integrating geographic information into the decision-making chain, achieving integrated spatial reasoning and action that “sees the map, understands the terrain, and selects positions accurately” [5].

In military simulation and battlefield decision support, selecting advantageous positions remains a core challenge [6]. Traditional rule-driven or reinforcement learning methods are often constrained by extensive scenario priors and manual feature designs, making generalizations to unseen terrains problematic. Recent efforts, such as COA-GPT [7], have used LLMs to assist in tactical deductions through natural language reasoning and tool use. However, much of the current research focuses on command planning, lacking fine-grained understanding and point selection capabilities for specific geographic locations.

Moreover, the emergence of visual large language models (VLLMs) has endowed LLMs with preliminary capabilities in map and scene understanding. Works like MapLM [8] and GeoGPT [9] have begun integrating geographic images and text descriptions for spatial localization and navigation. Additionally, SpatialVLM [10] trained VLM on large-scale 3D spatial reasoning datasets, enhancing capabilities in qualitative and quantitative spatial question answering. VLMMap [11] further unified visual features and semantic labels into a coordinated space for robot path planning and geographic command execution. However, most of these models focus on static recognition and localization tasks, lacking real-time reasoning capabilities for dynamic changes in geographic conditions, such as location alterations and visibility range modifications, while failing to develop an interpretable logic chain for point selection.

In AI gaming, point selection in games like Go and Chess—with agents such as AlphaGo, KataGo, and Leela Zero—offers insights into dynamic spatial decision-making. The “move selection” in Go essentially constitutes a two-dimensional spatial reasoning task, where AI evaluates and predicts spatial conditions via policy and value network [12, 13]. These systems can identify key positions within vast state spaces and deduce long-term benefits. However, compared to geographical point selection, the spatial rules in board games are structured and fully observable, lacking the uncertainty and geographic semantic information inherent in real-world scenarios, and direct transitions to military simulations are challenging.

In summary, critical gaps persist between existing research on language reasoning—represented by ReAct-like agents—and spatial understanding exemplified by VLLMs. Presently, no unified framework can concurrently comprehend geometric map structures, perceive geographic semantic information, and engage in reasoning and decision-making in dynamic simulation environments.

To address these gaps, this paper proposes the Geo-Commander framework, a multi-task LLM commander agent that combines the ReAct reasoning mechanism with spatial awareness. By encoding the terrain using a

hexagonal grid, Geo-Commander features a Geo-Choice module for preliminary screening of the hexagonal grid map, enabling the agent to establish spatial constraints early in the reasoning chain. The ReAct mechanism then incorporates geographic information into the language reasoning process, producing interpretable point selection decisions, as depicted in Figure 1. This study validates Geo-Commander through both static and dynamic experiments in a tank detachment combat simulation, covering three types of static point selection tasks (concealment, sniper, and assault points) and real-time dynamic command tasks. Results indicate that Geo-Commander significantly outperforms control models in terms of selection quality, win rate, and combat score.

The main contributions of this paper include:

1. Introducing an intelligent decision-making framework that integrates structured spatial encoding with language reasoning mechanisms, addressing the core limitations of traditional VLLMs in dynamic geospatial point selection.

2. The Geo-Choice module implements a semantic-based spatial coarse screening mechanism using hexagonal grids, reducing decision space complexity and providing an efficient reasoning solution in large geographic areas.

3. Designing a Spatialized ReAct Chain that embeds geographic semantic information into the VLLM's reasoning loop, forming a complete "Reasoning-Action-Observation" cycle, significantly enhancing decision interpretability and reliability in dynamic environments.

4. Validating the framework's superiority in tank detachment combat simulation through a multi-task evaluation system, significantly exceeding existing state-of-the-art methods across various metrics, setting a new benchmark for intelligent decision-making systems in complex environments.

5. Establishing a new research direction of "geographic intelligence + multi-agent collaboration," laying groundwork for advanced battlefield decision support systems and related areas.

Figure 1. Workflow of the Geo-Commander Framework. This diagram illustrates the mechanism of the Geo-Commander system. Geo-Commander converts real or virtual terrain into a hexagonal grid map through spatial state representation. It inputs both the hexagonal grid map in image form and spatial encoding into VLLM #1 in the Geo-Choice module. VLLM #1 matches task features supported by the hexagonal grid map, outputting spatial coarse screening results to VLLM #2 in the Spatialized ReAct Chain. VLLM #2, based on the hexagonal grid map image, analyzes the spatial screening results grid point by grid point, ultimately yielding the optimal task solution, which is fed back to the hexagonal grid map scenario.

2 Methodology

This chapter elaborates on the implementation details and technical

architecture of Geo-Commander. The Geo-Commander framework achieves precise spatial representation through hexagonal grid map encoding. It utilizes multiple VLLMs with visual understanding capabilities as the core of reasoning. Preliminary spatial state screening is completed through the Geo-Choice module, and realizes closed-loop decision-making of reasoning-action-observation with the help of the ReAct mechanism.

2.1 Spatial State Representation

In geospatial reasoning tasks, the form of map representation directly determines the model's ability to understand spatial relationships. Research indicates [5] that basic VLLMs struggle to understand spatial relationships because they cannot accurately correspond image content to textual semantics, whereas basic LLMs, although capable of understanding spatially structured inputs, base their overall understanding of space on pre-training and guessing, lacking an intuitive understanding of the space as a whole. To resolve this contradiction, this study adopts a hexagonal grid map as the fundamental spatial carrier, combined with the visual input of VLLMs, with the aim of equipping Geo-Commander with a structured cognition of the battlefield space. Hexagonal grid maps have advantages such as uniform adjacency relationships, continuous directionality, and simple distance calculation [14] and are widely used in military and non-military wargaming, combat simulation, and even board games and video games. This is possibly because the LLM pre-training corpus already contains a large amount of content related to hexagonal grid maps. In our preliminary tests, various LLMs and VLLMs demonstrated a profound understanding of the arrangement rules of hexagonal grid maps without requiring additional explanation, which provides a good foundation for our subsequent work.

In this framework, each grid point is defined as a spatial state unit $S_{i,j}$, and its encoding consists of the following four core dimensions.

(a) Column coordinate (Column, q) and row coordinate (Row, r): An "offset coordinate system" is used to describe the grid point position, ensuring consistent adjacency relationships in all six directions.

$$S_{i,j} = (x_i, y_j) \quad (1)$$

where x_i represents the column coordinate and y_j represents the row coordinate. The adoption of this hexagonal grid spacing is determined by the granularity of the combat simulation. The tank detachment combat simulation supported by this study actually details combat at the platoon-granular. Following conventional practice in this field [15], to balance the spatial resolution and abstraction, a 200-meter distance between adjacent grids was adopted, which is sufficiently accurate for describing platoon-granular combat in various terrains.

(b) Elevation: Reflects terrain relief and line-of-sight occlusion features. To ensure universality and a certain level of granularity, we represent the grid point elevation by averaging the altitude within the grid point area as a

multiple of 5, ensuring that the LLM can relatively easily identify "high ground" and "low ground" relationships through comparison.

(c) Geographical Elements: Represented by symbolic codes such as Open Ground (P), forest (F), building (B), and river (R), etc. Each grid point contains one geographical element. To facilitate LLM understanding, we emphasized the geographical elements corresponding to the symbolic codes in the prompts.

Finally, the entire map can be regarded as a two-dimensional tensor:

$$M = \{S_{i,j} | i \in [1,W], j \in [1,H]\} \quad (2)$$

Where W and H are the map width and height, respectively.

Before inputting the hexagonal grid map screenshot into the model, the coordinates and elevation were directly marked as numerical values inside each hexagon, while different geographical elements were marked as icons inside each hexagon. All screenshots in this study were presented at 144 dpi resolution using 18pt Arial digits for coordinates and elevation. Legends provide semantic anchors for VLLMs, ensuring the alignment of map symbols with the text [16, 17]. We included a legend outside the screenshot to explain the icons corresponding to different geographical elements.

2.2 LLM Selection and Adaptation

The core reasoning engine of Geo-Commander is based on a Visual Large Language Model (VLLM) with multimodal understanding capabilities to achieve joint spatial perception of images and text. Model selection followed the following principles:

(a) It possesses an image input channel and has a correct perception of the input hexagonal grid map. In other words, the model can directly receive screenshots of the hexagonal grid map and describe terrain features in a natural language form. When inputting a basic hexagonal grid map that does not yet contain combat units and legends, we ask, "Can you see the uploaded image?" General VLLM models can usually output correct summaries, such as "This is a hexagonal grid map of a map, which looks like a strategy game or geographic information system map. The map is mainly composed of many hexagonal grids, each marked with numbers, possible coordinates, or numbers. The background color of the map is light yellow and the borders of the hexagonal grids are dark yellow. On the right side of the map, there is a green area, possibly representing forest or vegetation.", and other semantic summaries. Based on this method, we also tested some professional geographic VLLMs, but the results were unsatisfactory, such as "This geological map shows a detailed regional geological structure, including different strata, faults, and geological features. The following is a detailed description..." This may be because these geographic VLLMs are trained to be professional geological tools. VLLMs that could not correctly describe the hexagonal grid map were excluded.

(b) Support tool calling and intermediate-reasoning chain generation. To implement the ReAct mechanism, the model needs to be compatible with the multi-turn closed-loop interaction capability of "reasoning-action-observation-re-reasoning." Currently, most VLLMs satisfy this requirement.

(c) It possesses a strong context memory window. Because point selection reasoning often involves multi-turn map state analysis and tool calls, long context support can significantly reduce failure rates and improve accuracy. This method excludes small-scale VLLMs.

Combined with the VLLM benchmark evaluation results [18] and while retaining the largest and most effective VLLM within the same series of VLLMs, we pre-selected four VLLMs for the experiment. Furthermore, we conducted a targeted "Visual-Textual Alignment Test" on the pre-selected VLLMs, which involved uploading a screenshot of the hexagonal grid map and the legend to the VLLM, and asking "Please describe the terrain features at coordinate XX in the image", by observing whether the VLLM can correctly identify geographical elements and elevation. Finally, we compared the effectiveness of different VLLMs in the "Visual-Textual Alignment Test." The results of the "Visual-Textual Alignment Test" are shown in Figure 2.

As shown in Figure 2, even the most effective VLLM did not achieve perfect recognition accuracy. Such errors can prevent the VLLM from selecting points entirely according to the task requirements. Therefore, measures are required to ensure that the VLLM accurately understands the map and prevents errors.

Figure 2. Visual-Textual Alignment Test Results. This test is based on the hexagonal grid map screenshot and legend we provided, selecting ten grid point coordinates from the hexagonal grid map, and asking questions coordinate by coordinate using a zero-shot method. Qwen-VL-Max and GPT-4V showed relatively accurate understanding of the hexagonal grid map. Qwen-VL-Max could accurately identify the information contained within the hexagons in most cases; GPT-4V seemed unable to understand the geographical elements in the hexagons; ERNIE-4.5-Turbo-VL mistakenly regarded coordinates as elevation and could not be corrected by modifying the legend, resulting in zero accuracy in elevation recognition; DeepSeek-VL2 over-focused on the legend, did not carefully observe the map, and could not be corrected by modifying the legend, resulting in zero accuracy in elevation recognition.

2.3 Geo-Choice Module

The Geo-Choice module is a key intermediate layer in Geo-Commander that connects map perception with language reasoning. Its role is to enhance the visual description consistency of the intelligent agent and perform "spatial coarse screening" on the map space before ReAct reasoning, thereby

reducing the original map state space from potentially thousands of grid points to a limited candidate set $C=\{c_1, c_2, \dots, c_n\}$. To balance the scope and accuracy of spatial coarse screening, we set the candidate set to $n \leq 10$.

The module design is divided into two steps:

(a) Text Input for Spatial Encoding. This module uses an automatic script to convert each grid point of the hexagonal grid map into a dictionary format, for example, "(4,7), 105, P," and inputs it as text to VLLM #1 within the module. This structured map representation preserves the spatial structure while reducing the number of input characters, enabling the VLLM to achieve text-image alignment and ensure that it accurately understands the map. By adopting Text Input for Spatial Encoding, we have effectively eliminated the issue of misreading in VLLM as much as possible. To illustrate this, we conducted a "Misreading Elimination Test," which showed that after providing textual coordinate input, the understanding ability of all tested VLLMs regarding the hexagonal grid map significantly improved (see Figure 3). This enhances the accuracy of Geo-Commander's decision-making.

Figure 3. Results of the Misreading Elimination Test. This experiment is the Visual Textual Alignment Test conducted after Text Input for Spatial Encoding to VLLM. Ten grid point coordinates were chosen from the hexagonal grid map, and questions were posed coordinate by coordinate, utilizing a zero-shot approach. After providing Text Input for Spatial Encoding, the abilities of all tested VLLMs were improved. Qwen-VL Max and GPT-4V answered all the questions correctly in the test, and ERNIE-4.5-Turbo-VL and DeepSeeker VL2 also improved their understanding of hexagonal grid map.

(b) Task Feature Matching. Based on conventional army tactics, land warfare is divided into three phases [19]: Assembly (Preparation), Firepower, and Melee Phase. The Assembly Phase task is to concentrate forces, requiring the selection of concealment points that are far from the enemy and not easily observed; the Firepower Phase task is to snipe targets from a long distance, requiring the selection of sniper points with good visibility and concealment; and the Melee Phase task is to approach or break into the enemy formation for close-quarters combat, requiring the selection of assault points that are close to the enemy and conducive to maneuvering. According to military common sense, it is easy to see that the selection requirements for concealment points also make them suitable for retreat and evasion tasks, and the selection requirements for sniper points make them suitable for defense and covering tasks [20, 21]. Furthermore, specific rules regarding how geographical elements and elevation cause spatial occlusion and how they affect maneuverability and concealment need to be provided. These knowledge provide clear optimization criteria for the Geo-Choice module. Based on this task feature matching knowledge, the Geo-Choice module first matches the tactical task type based on the task requirement text, and then

preliminarily screens a set of candidate grid points that may meet the requirements according to the current tactical task.

The overall mechanism of the Geo-Choice module is depicted in Figure 4. This module outputs a candidate set of adjacent or non-adjacent cells for the downstream ReAct module to select, and these cells will be independently and discretely measured. In all our tests, this module has never output duplicate cells, has demonstrated stable performance.

The above design enables the intelligent agent to perform text-image alignment, and instead of searching the entire map, it focuses on deep reasoning within a limited, semantically interpretable spatial subset, significantly improving accuracy and reducing context length, computation, and token consumption costs.

Figure 4. Mechanism of the Geo-Choice Module. The screenshot of the hexagonal grid map, spatial encoding, and task feature matching knowledge are input into the VLLM of this module. The VLLM provides the candidate set based on the input information.

2.4 Spatialized ReAct Chain

The ReAct (Reasoning+Action) mechanism is the core of the Geo-Commander's spatial reasoning. Its principle lies in alternately executing the two stages of "language reasoning" and "action operation," with each step being corrected based on environmental feedback (Observation), thus forming an interpretable reasoning closed loop [1].

In the Geo-Commander, the ReAct mechanism is extended to a spatialized version (Spatialized ReAct Chain). The reasoning process is as follows.

(a) Reasoning Stage: VLLM #2 analyzes the current task and terrain semantics, and generates preliminary hypotheses based on the candidate grid points provided by the Geo-Choice module. For example, "the current task is defence, requiring the selection of a sniper point with good visibility and concealment. The enemy is from (25,47). Based on preliminary screening, (28,52), (28,54), (25,54)... may meet task requirements. Each location needs to be analyzed specifically to determine the best sniper point among them."

(b) Action Stage: VLLM #2 calls the map tool (MapTool) to perform specific operations, such as "calculate the line-of-sight area of a certain grid point," "calculate the cost for our unit to maneuver to a certain grid point," "calculate the distance between a certain grid point and the enemy unit." The measurements made by MapTool for each grid point are independent and discrete. MapTool provides the following functional interfaces:

get_visibility(coord): Input grid point coordinates, return all line-of-sight grid points, and the number of line-of-sight grid points for that coordinate.

get_maneuver_cost(coord): Input the grid point coordinate and return the time required for our nearest unit to maneuver to that coordinate.

get_distance_enemy(coord): The input grid point coordinate returns the

distance from the coordinate to the nearest enemy unit.

The specific calculation rules for each function in MapTool are aligned with the simulation environment that supports the operation of hexagonal grid maps. In practice, the line-of-sight calculation performs a simple traversal within a 25-cell distance, judging each traversed cell as yes or no; each tool call takes only a few milliseconds. The maneuver cost calculation employs a customized A* algorithm; each tool call takes an average of 10-20 milliseconds. The distance calculation is a point-to-point arithmetic operation; each tool call takes less than 1 millisecond. In practice, after each tool is called 10 times, the total time consumed does not exceed 1 second.

(c) Observation Stage: The system returns the function results, and VLLM #2 updates its internal state accordingly and proceeds to the next reasoning cycle or terminates the cycle based on the task requirements and candidate grid points.

(d) Answer Stage: VLLM #2, based on the data obtained from the Reasoning-Action-Observation loop, compares and analyzes the best position that meets the current task requirements and outputs the final decision. To exert the emergent and adaptive capabilities of LLM, we do not provide restrictive appraisal rules but only appraisal principles. For example, "(25,54) has the most line-of-sight area covering potential enemy positions and possesses concealment (the geographical element is building). Choose (25,54) as the sniper point."

To control the reasoning complexity, the system sets the maximum number of cycles to $N=10$. If VLLM #2 has not entered the Answer stage by the 10th cycle, an automatic summary mechanism is triggered, and VLLM #2 gives the final result based on the existing intermediate results. This explicit reasoning chain provides traceable decision logic, significantly enhancing the interpretability of intelligent agents. The overall mechanism of the Spatialized ReAct Chain is depicted in Figure 5.

Figure 5. Mechanism of the Spatialized ReAct Chain. The current tactical task requirements and the coordinate candidate set are input as initial information into the VLLM of this module. The VLLM performs point-by-point measurements, compares and analyzes the optimal position through the Reasoning—Action—Observation loop, and finally outputs the command.

2.5 Ethics statement

Ethical approval was not applicable for this study. The involvement of the military experts in this study falls under the category of professional consultation and expert evaluation. The experts provided quality ratings for the established tank combat simulation scenarios based on their professional knowledge and experience. This process did not involve the collection of any personal sensitive information, medical data, or any form of intervention. In accordance with the spirit of China's "Ethical Review Measures for Life Science

and Medical Research Involving Humans" and common academic practice, such expert consultations are typically eligible for exemption from ethical review. Therefore, a separate IRB approval was not sought for this specific part of the study. Additionally, the Ethics Committee of the Academy of Military Medical Sciences has confirmed that ethical approval was not required for this study.

This study adhered to all relevant national regulations and international ethical standards, including the Declaration of Helsinki. All participants (the four military experts) employed in this study were adults. Before their participation, all individuals were fully informed about the purpose of the study, the procedures involved, and how their input would be used. Written informed consent was obtained from all participants before their involvement in the study.

3 Experimental Validation

To verify the effectiveness of the Geo-Commander framework in dynamic geographic point selection and tactical command tasks, this study designed a systematic experimental plan. The experiments were conducted at two levels: static spatial reasoning ability and dynamic combat simulation task performance, covering multi-dimensional indicator evaluation and multiple ablation control groups, to comprehensively evaluate the performance of intelligent agents under different spatial topologies and task types.

3.1 Experimental Environment and Scenarios

The experiments were conducted based on an Army Tactical Intelligent Wargaming Simulation Platform [22] using B/S architecture. This simulation environment is specifically used for army combat simulations and employs engagement adjudication rules based on random numbers and adjudication tables. Similar to common army combat simulations, terrain affects the unit line-of-sight and maneuver speed, and various factors such as geographical elements, the unit's own state, and target distance affect the engagement adjudication results. Apart from the basic maneuver rules, reconnaissance rules, and engagement rules, this simulation environment has no other special mechanisms. We chose this simulation environment because the Army Tactical Intelligent Wargaming Simulation Platform uses a time-step-based interaction triggering mechanism. The platform's time step does not advance until all intelligent agents have completed their decisions. This mechanism ensures that the agent's capability is not affected or limited by the decision time, making it suitable for testing LLM-based agents.

Based on this simulation environment, we set up multiple hexagonal grid maps using the spatial state representation method. The overall map size was set from 13×23 grids (approximately 12 km^2) to 100×100 grids (approximately 400 km^2), containing various geographical elements, such as open ground, forest, river, and building areas. The elevation and geographical

elements on the map were generated by a program based on real satellite maps and manually corrected to ensure reasonable terrain and a certain level of spatial complexity. The experimental scenarios are illustrated in Figure 6. Terrains a,b,c, and d are sequentially: "Urban Residential Area"(size: 90×90) with widely distributed buildings and roads; "Paddy Field Area"(size: 80×80) where soft ground is everywhere causing significant reduction in tank maneuver speed; "Plateau Corridor"(size: 100×100) with steep terrain and few shelters; "Mountainous and Jungle Area"(size: 13×23) with undulating terrain and dense forests. The initial force deployment was configured according to the conventional tank detachment. Each side's tank detachment consisted of three platoon units. To simplify unrelated conditions, the initial distance between the two tank detachments was set to 10-20 grids (2-4 km) to ensure that the central terrain blocks the initial line-of-sight for both sides, and the central area was set as a control point to provide confrontation conditions.

Figure 6. Experimental Scenarios for Tank Detachment Combat Simulation. To maneuver the red tank platoon against the blue force, the intelligent agent under test must specify target coordinates. Each tank platoon has a maximum attack range of 15 grids and can automatically detect targets within 25 grids in the absence of terrain obstacles. The firing rule for both sides' tank platoons is to automatically attack the nearest target within line-of-sight.

3.2 Experimental Setup

Based on the experimental environment and scenarios, the intelligent agent to be tested is responsible for selecting coordinates to command the maneuver of the red tank platoon units and opposing the blue side. The Blue side is set as a rule-based commander intelligent agent "DEMO_AI" with a certain level of capability within the simulation platform. DEMO_AI has dynamic adjustment capabilities and multiple strategy rules.

Ablation experiments can effectively reveal the contribution of the Geo-Choice module to spatial constraints and task relevance, the gain of the ReAct mechanism in reasoning consistency and interpretability, and the impact of their synergy on task performance and token efficiency. To analyze the role of each module in the overall performance and compare representative designs in this field, this study sets up the following types of control agents:

Geo-Commander: The complete agent containing the Geo-Choice module and Spatialized ReAct Chain.

w/o ReAct: A Geo-Commander agent that retains the Geo-Choice module but removes the Spatialized ReAct Chain.

w/o Geo-Choice: A Geo-Commander agent that retains the Spatialized ReAct Chain but removes the Geo-Choice module.

Baseline: A Geo-Commander agent with both the Geo-Choice module and

the Spatialized ReAct Chain removed, performing point selection based only on the hexagonal grid map and task requirements, which can also represent traditional VLLMs.

DEMO_AI: A representative rule-based commander intelligent agent that deeply accesses the underlying architecture of the simulation platform for terrain information acquisition, and is the same as the agent used by the blue side. The introduction of this agent helps illustrate the balance of the scenario for both Red and Blue. Only applicable to dynamic experiments.

AlphaWar[23]: A representative reinforcement-learning-based commander intelligent agent specifically designed for tank detachment combat simulation. It utilizes professional commander duel experience for policy network pre-training and then undergoes reinforcement learning iterations through 1M self-play episodes. Its performance has reached the state-of-the-art (SOTA) level in this field, and to our knowledge, it is also the agent with the highest score and winning rate among the currently publicly available intelligent agents in this dynamic experiment. It is a fully trained out-of-the-box agent. It accesses the underlying architecture of the simulation platform for terrain information acquisition. Only applicable to dynamic experiments.

According to preliminary testing, none of the existing VLLM intelligent agents specifically enhanced for geography or spatial tasks are capable of performing the point selection decision-making task for a combat simulation commander (some have mismatched task types and output irrelevant results, whereas others, although able to output as required, perform worse than traditional VLLMs). Therefore, they were not included as controls. For all the VLLMs used in the experiments, we set the context window to 128 K and the temperature value to 0.1.

The experimental tasks consisted of both static and dynamic experiments. This study uses the series of static tests on Terrain-a as an example to verify the performance superiority of different VLLM agents in each basic point selection task, serving as a supplementary evaluation of the VLLM agents' point selection capabilities at various stages in the dynamic experiment. The static experiments consisted of three basic point selection tasks: selecting concealment points (to avoid attacks and force assembly), selecting sniper points (for defense and ambush), and selecting assault points (for offense). The experimental metric used was an expert-designed evaluation standard. Four military experts - including an army commander, an army staff officer, and two wargaming experts - were employed to analyze the experimental scenarios, incorporating all the aspects that need to be considered for the tank combat simulation point selection tasks, to ensure a comprehensive and profound understanding of this task. Through discussion, the four military experts provided a grid point quality rating table for the three basic point selection tasks in the experimental scenarios, as shown in Table 1. During the discussion, the experts comprehensively analyzed the characteristics of each grid cell and reached a consensus through full discussion, forming highly

consistent opinions ($\kappa \approx 0.95$). During each test, if the output result of the intelligent agent corresponds to the "optimal" rating in the grid point quality rating table, it scores 1 point; if it corresponds to the "suboptimal" rating, it scores 0.5 points; otherwise, it scored no points. Each task in the static experiment was run 100 times and a percentage score was given. This rating table is intended to serve as a performance evaluation benchmark for each agent on the basic point selection tasks in the Terrain-a static experiment and is only applicable for scoring in the Terrain-a static experiment.

Table 1. Grid Point Quality Rating Table (Terrain-a)

	Optimal	Suboptimal
Concealment Point	(24,56); (25,55)	(21,54); (21,55); (21,56); (22,55); (23,54); (23,55); (24,55)
Sniper Point	(23,52); (25,54); (26,52); (28,52)	(26,53); (23,53); (24,52); (24,53)
Assault Point	(24,52); (25,54); (27,53)	(25,53); (26,52); (27,52)

In the dynamic experiment, the basic point selection tasks from the static experiment were embedded into a standard tank detachment combat simulation scenario. The intelligent agent commands the tank battle through a phased point selection throughout the game. That is, according to conventional land warfare tactics, select a concealment point in the assembly phase at the beginning of the game, select a sniper point in the firepower phase after completing assembly, select an assault point in the melee phase when enemy units attack the "main point," and after destroying all targets, sequentially occupy the "main point" and the "minor point." The experimental metrics were the net score and win rate. In the Army Tactical Intelligent Wargaming Simulation Platform, the net score calculation rule is as follows: the net score is 0 at the start; for each destroyed enemy tank platoon (composed of two tanks with a value of 20 each), 40 points were added to one's own side; for each loss of one's own tank platoon, 40 points are deducted from one's own side; at the end of the game, occupying one "minor point" adds 50 points, and occupying one "main point" adds 80 points. A net score greater than zero was considered a win. Each test in the dynamic experiment was repeated 100 times. Finally, the average net score and win rate were provided. The dynamic experiments were conducted in a locally deployed simulation environment. All agents (including Geo-Commander, AlphaWar, DEMO_AI, etc.) interacted with the simulation platform through the same local interface, completely eliminating network latency effects. All agents were run in the exact same simulation environment, with the same global observation space, the same action space, and under conditions guaranteeing sufficient and complete computational resources, ensuring that performance comparisons depended solely on the agent algorithms

themselves.

3.3 Experimental Results

Static experiment testing was conducted in Terrain-a, and the results are shown in Figure 7.

It can be found that, based on our quite strict grid point quality rating table, traditional VLLMs have no advantage in point selection tasks, and find it difficult to correctly understand and analyze space. Overall, Geo-Commander with GPT-4V as the cognitive core achieved the best comprehensive effect; ERNIE-4.5-Turbo-VL and Qwen-VL-Max also showed good comprehensive effects, whereas DeepSeek-VL2's comprehensive effect was relatively poor. Although Qwen-VL-Max won in the visual description consistency test, the two major gain module designs in this study significantly compensated for the shortcomings of GPT-4V and ERNIE-4.5-Turbo-VL in visual description consistency, allowing their intelligence to be fully utilized and enabling them to surpass Qwen-VL-Max in point selection quality score. The poor performance of DeepSeek-VL2 might be due to its relatively small number of parameters limiting its intelligence ceiling. Even so, the two major gain module designs in this study still improved the score of the agent using DeepSeek-VL2 as the cognitive core.

From the perspective of point selection task scores, the scores for concealment points are generally low, possibly because the selection range for concealment points is large, and the VLLM only analyzes based on its own understanding of the map and the results of map tool calls, lacking sufficient understanding of terrains like "reverse slopes" that require certain professional knowledge and rely more on intuitive judgment.

From the perspective of ablation setup scores, by comparing with Baseline, it can be seen that the Geo-Choice module is suitable for finding relatively suitable grid points from point selection tasks with a large selection range (concealment points and sniper points), but its advantage is not obvious in point selection tasks with a small selection range (assault points), which is very consistent with the design characteristics of this module. Similarly, it can be found that the Spatialized ReAct Chain has obvious advantages in point selection tasks with a small selection range (assault points), but its advantages are smaller in point selection tasks with a large selection range (concealment points and sniper points) because the Spatialized ReAct Chain can only select a limited number of grid points for analysis, which heavily relies on the point selection range provided by the previous process. There are also exceptions here. For example, the effect of GPT-4V's Spatialized ReAct Chain on the sniper point selection task is significantly better than its effect on the assault point selection task, which may be because GPT-4V has a stronger intuitive ability in the sniper point selection task, providing a relatively reasonable point selection range for the Spatialized ReAct Chain, and the source of this particularity may be GPT-4V's unique architecture and

pre-training method.

Figure 7. Heatmap of Static Experiment Point Selection Quality Score. Four VLLMs were tested. Different ablation settings are shown by the vertical axis, while various point selection tasks are represented by the horizontal axis. Higher scores are shown by darker colors on the 0–100 point selection quality score scale.

Figure 8. Results of Ablation Dynamic Experiments. The cognitive core for Geo-Commander was chosen to be ERNIE-4.5-Turbo-VL, which was evaluated on Terrain-a. The average score and win rate are shown on the vertical axis, while various intelligent agent parameters are represented on the horizontal axis.

Figure 9. Results of Generalization Dynamic Experiments. The cognitive core for Geo-Commander, ERNIE-4.5-Turbo-VL, was chosen and tested across various terrains. The win rate and average score are shown accordingly.

For ablation comparison and generalization testing in dynamic experiments, we selected ERNIE-4.5-Turbo-VL as the cognitive core for Geo-Commander because of its representative score distribution in the static experiments — high Geo-Commander score, medium module scores, and not too low a baseline score. The results are shown in Figure 8 and 9, respectively.

In the ablation dynamic experiments, a win rate of DEMO_AI close to 50% indicated that the scenario was fair for both combatants. From the ablation settings perspective, the baseline without any additional modules shows worse performance than DEMO_AI, whereas adding either module results in a slightly better performance than DEMO_AI, but worse than the previous SOTA agent AlphaWar. Geo-Commander, possessing both modules, has an average score and win rate that significantly exceeds that of the previous SOTA agent AlphaWar, showing very strong performance, indicating that the excellent performance of Geo-Commander comes from the combined effect of the Geo-Choice module and the Spatialized ReAct Chain.

In the generalization dynamic experiments, different scenarios had some influence on the Geo-Commander's average score and win rate. Compared to Terrain-a, Terrain-b has more types of geographical elements, complex distribution, and a larger combat area, which poses greater challenges for Geo-Commander; therefore, the improvement level in the average score and win rate is lower than in Terrain-a. Even so, the improvement is still very significant. Terrain-c has fewer types of geographical elements and a simple distribution; here, Geo-Commander can complete correct point selection, and the average score and win rate are greatly improved. Terrain-d has geographical element types and distribution comparable to Terrain-a, but the

combat area is smaller, which enhances the Geo-Commander's performance level. Overall, the Geo-Commander achieved SOTA levels in multiple scenarios, demonstrating its good generalization ability.

We further supplemented the generalization test with GPT-4V as the cognitive core of the Geo-Commander, and the results are shown in Figure 10. Overall, the performance was better than that of ERNIE-4.5-Turbo-VL. Surprisingly, GPT-4V seemed to possess unique resilience, achieving higher scores in the relatively complex terrains of Terrain-a and Terrain-b. The source of this particularity might be the large parameter size of the GPT-4V.

Figure 10. Results of the Generalization Dynamic Experiment using GPT-4V as Geo-Commander's Cognitive Core. Every terrain was used for the experiments. The win rate and average score are shown accordingly.

4. Discussion

The Geo-Commander framework focuses on enhancing dynamic geographic point selection in combat simulations through a novel integration of spatial encoding and reasoning systems. Although decisions are made infrequently in practical problems, token usage and response speed must still be considered to meet the basic decision frequency requirements. Presently, a single run of the Geo-Commander consumes 3000-3500 tokens and takes 40-50 seconds, of which the Geo-Choice module typically consumes 800-1000 tokens and takes 20-25 seconds, and the Spatialized ReAct reasoning chain typically consumes 2000-2500 tokens and takes 20-25 seconds. For combat command at platoon-granular, a decision cycle within one minute is acceptable. Many platoon-granular combat simulation models, wargaming systems, and ground combat simulations adopt a minute-scale resolution as their fundamental time step, as seen in systems such as PanzerBlitz, Panzer '44, and A Fistful of TOWs. In actual combat command, if one wishes to further improve decision speed, different encounter scenarios can be created in peacetime exercises. Geo-Commander can then automatically determine the best deployment location for these scenarios, which can then be stored as backup plans following manual review and augmentation, for rapid invocation. Overall, the token usage and response speed of the Geo-Commander fully meet the requirements of practical applications.

The experimental results demonstrate that by integrating structured spatial encoding and an effective reasoning mechanism, Geo-Commander shows a robust spatial understanding in complex geographic contexts. Geo-Commander not only maintains logical coherence at the language level but also exhibits reasonable spatial constraints at the geographic semantic level. This indicates a significant complementary relationship between language reasoning and spatial perception, and that the multimodal capabilities of LLMs can be fully utilized through structured spatial input and map tool calls. This synergy between language reasoning and spatial

perception significantly enhances the interpretability and reliability of the decision-making process, paving the way for applications beyond military simulations, such as disaster response and automated planning.

It is worth noting that the general reasoning capability and image recognition accuracy of the underlying VLLM will affect the performance of the Geo-Commander to some extent, which is an unavoidable limitation for all intelligent agents using LLM as the cognitive core. The poor performance of DeepSeek-VL2 indicates that if the parameter size of the underlying VLLM is too small, it will severely limit the upper limit of the Geo-Commander's capabilities, and using a VLLM with a larger parameter size would yield better results, such as GPT-4V. Moreover, when the input map scale is too large, the number of geographic elements is too high, or the terrain features are too dense, the model may fail to identify optimal grid points during preliminary screening, leading the reasoning chain to select from suboptimal options.

The task feature matching rules in the Geo-Choice module are based on the three-phase division of traditional army tactics, as these are well-established and provide clear point selection requirements, enabling Geo-Commander to fully utilize its dynamic spatial reasoning capabilities. Emerging tactical tasks that are not pre-defined, such as unmanned combat, are currently undergoing rapid development and iteration, and their tactical theories are not yet solidified. Therefore, it is also difficult to provide Geo-Commander with clear point selection requirements. Although Geo-Commander is currently based on traditional army tactics, its design allows for knowledge base updates and has the potential to adapt to emerging tactical tasks. As military theories evolve and become more refined in the future, corresponding upgrades can be made.

It should be noted that the dynamic experiments in this study are primarily based on traditional tank detachment combat scenarios and have not yet covered more complex practical situations such as asymmetric warfare and multi-service coordination. Nevertheless, the design of Geo-Commander allows for knowledge base updates and possesses the potential to adapt to more complex combat situations. Future research will focus on validating and enhancing its generalization capabilities in more complex combat environments.

5. Conclusion

This study introduces the Geo-Commander framework as a groundbreaking approach to applying LLMs in dynamic spatial reasoning tasks. Through the development and integration of the Geo-Choice module along with the Spatialized ReAct Chain, our experimental findings reveal significant improvements in both static point selection tasks and real-time command tasks in combat simulations.

The validation of Geo-Commander demonstrates the effective integration of language reasoning and spatial understanding, providing a viable path to

transition LLM capabilities from traditional linguistic tasks to complex spatial intelligence. This framework lays foundational groundwork for future research in multi-agent collaboration and higher-level tactical decision-making in dynamic environments.

Geo-Commander's study aimed to explore the potential of AI rather than replacing human command responsibilities. To avoid the risk of misuse of autonomous weapon systems, this study was strictly limited to a closed, non-real combat simulation environment for training, analysis, and tactical deduction. After further development of the HCI interface, all decision outputs of the system should be considered as auxiliary recommendations and must be reviewed, modified, and finally approved by the human commander before implementation. One of the core purposes of introducing reasoning mechanisms and modules into this framework is to improve the interpretability of the decision-making process. The system needs to state its reasoning basis step by step, which enables the human commander to track, understand, and question its decision logic to effectively supervise.

Acknowledgements: We would like to express our sincere gratitude to the military experts who contributed their professional knowledge and time to this study. Specifically, we thank the Army Commander, the Army Staff Officer, and the two wargaming experts for their in-depth analysis of the combat simulation scenarios and for their crucial role in developing the quality rating table for grid points through discussion. Their expertise ensured the tactical relevance and validity of our experimental evaluation metrics.

Funding: This research received no external funding.

Conflict of interest: The authors declare no conflicts of interest.

Ethical Approval: Ethical approval was not applicable for this study. The involvement of the military experts in this study falls under the category of professional consultation and expert evaluation. The experts provided quality ratings for the established tank combat simulation scenarios based on their professional knowledge and experience. This process did not involve the collection of any personal sensitive information, medical data, or any form of intervention. In accordance with the spirit of China's "Ethical Review Measures for Life Science and Medical Research Involving Humans" and common academic practice, such expert consultations are typically eligible for exemption from ethical review. Therefore, a separate IRB approval was not sought for this specific part of the study. Additionally, the Ethics Committee of the Academy of Military Medical Sciences has confirmed that ethical approval was not required for this study.

Informed Consent: All participants (the four military experts) employed in this study were adults. Before their participation, all individuals were fully informed about the purpose of the study, the procedures involved, and how their input would be used. Written informed consent was obtained from all participants before their involvement in the study.

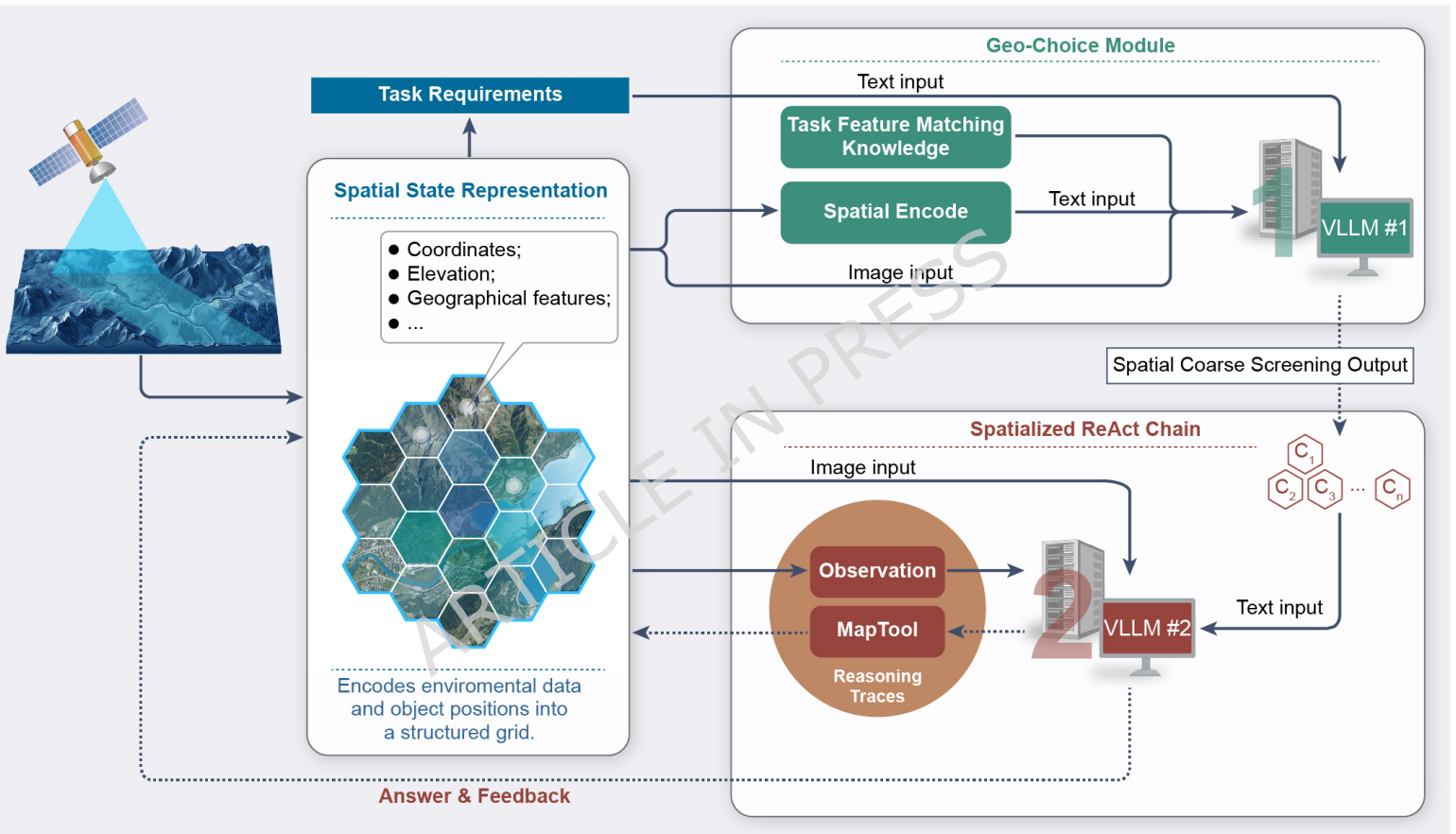
Author Contributions: Conceptualization, Yibo Chen; methodology, Caleb Jojo; software, Yibo Chen; validation, Shuhang Zhou; formal analysis, Shuhang Zhou; investigation, Caleb Jojo; resources, Yang Ping; data curation, Shuhang Zhou; writing—original draft preparation, Yibo Chen; writing—review and editing, Yang Ping; visualization, Caleb Jojo; supervision, Yang Ping; project administration, Yang Ping; funding acquisition, Yang Ping. All authors agree to be accountable for all aspects of the work.

Data Availability: The data that support the findings of this study are openly available in ScienceDB at <https://doi.org/10.57760/sciencedb.32513>, reference number [24].

Reference

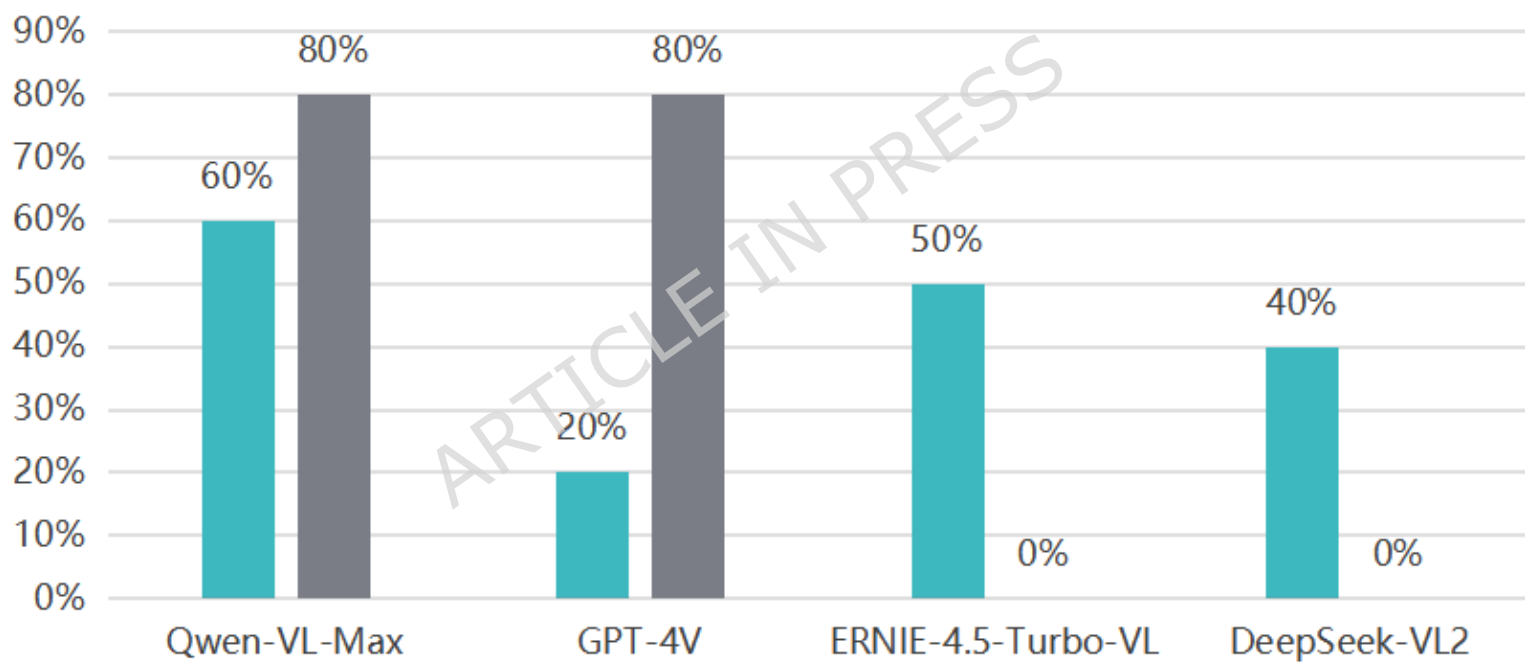
- [1] Yao S, Zhao J, Yu D, et al. REACT: SYNERGIZING REASONING AND ACTING IN LANGUAGE MODELS: 11th International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, 2023[C].
- [2] Shinn N, Cassano F, Gopinath A, et al. Reflexion: Language Agents with Verbal Reinforcement Learning: 37th Conference on Neural Information Processing Systems, NeurIPS 2023, New Orleans, 2023[C].
- [3] Wang G, Xie Y, Jiang Y, et al. VOYAGER: An Open-Ended Embodied Agent with Large Language Models[J]. arXiv, 2023.
- [4] Jafarnejad S. MapLLM - A Blueprint for Improving Geospatial Reasoning in LLMs[J]. 2025.
- [5] Cheng A, Yin H, Fu Y, et al. SpatialRGPT: Grounded Spatial Reasoning in Vision-Language Models: 38th Conference on Neural Information Processing Systems, NeurIPS 2024, Vancouver, BC, Canada, 2024[C].
- [6] Headquarters D O T A. FM3-34.230-Topographic Operations[M]. 2000.
- [7] Goecks V G, Waytowich N. COA-GPT: Generative Pre-Trained Transformers for Accelerated Course of Action Development in Military Operations: 2024 International Conference on Military Communication and Information Systems (ICMCIS), Koblenz, Germany, 2024[C].
- [8] Cao X, Zhou T, Ma Y, et al. MAPLM: A Real-World Large-Scale Vision-Language Benchmark for Map and Traffic Scene Understanding: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, United states, 2024[C].
- [9] Zhang Y, Wei C, Wu S, et al. GeoGPT: Understanding and Processing Geospatial Tasks through An Autonomous GPT [arXiv][J]. arXiv, 2023.

- [10]Chen B, Xu Z, Kirmani S, et al. SpatialVLM: Endowing Vision-Language Models with Spatial Reasoning Capabilities: 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, United states, 2024[C].
- [11]Huang C, Mees O, Zeng A, et al. Visual Language Maps for Robot Navigation: 2023 IEEE International Conference on Robotics and Automation (ICRA), London, United kingdom, 2023[C].
- [12]Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. *Nature*, 2016,Vol.529(No.7587):484-489.
- [13]Wu D J. Accelerating Self-Play Learning in Go[J]. *Statistics*, 2019,Vol.2.
- [14]Birch C P D C, Oom S P, Beecham J A. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology[J]. *Ecological Modelling*, 2007,Vol.206(NO.3-4):347-359.
- [15] Tang Fen,Zhang Xin,You Xiong,et al . Design Method of Tactical Level Hexagonal Wargame Map[J]. *Journal of System Simulation*, 2019, 31(5): 869-878.
- [16]Srivastava V, Lei F, Mukhopadhyay S, et al. MapIQ: Benchmarking Multimodal Large Language Models for Map Question Answering[J]. 2025.
- [17]R A I, P S, S V K E, et al. Light-MLLMAD: A Lightweight Multimodal Large Language Model for One-Shot Industrial Visual Anomaly Detection[J]. 2025.
- [18]Ruan J, Jiang D, Gao X, et al. MME-SCI: A Comprehensive and Challenging Science Benchmark for Multimodal Large Language Models[J]. 2025.
- [19]Headquarters D O T A. FM 3-21.8 - The Infantry Rifle Platoon and Squad[M]. 2007.
- [20]Headquarters D O T A. FM 3-06.11 - Camouflage, Concealment, and Cover[M]. 2002.
- [21]Headquarters D O T A. FM 3-05.222 / TC 23-14 - Sniper Training and Employment[M]. 2003.
- [22]Zhou, Z., Zhao, H.: Design and implementation of B/S architecture-based wargaming system. *Journal of Equipment Academy* 27(2), 68-72 (2016)
- [23]Yin, Q., Zhao, M., Ni, W., et al.: Intelligent Decision-Making Technologies and Challenges in Wargaming. *Acta Automatica Sinica* 49(5), 913-928 (2023)
- [24] Chen Y. Experimental Data of Geo-commander[DS/OL]. Science Data Bank, 2025. <https://doi.org/10.57760/sciencedb.32513>



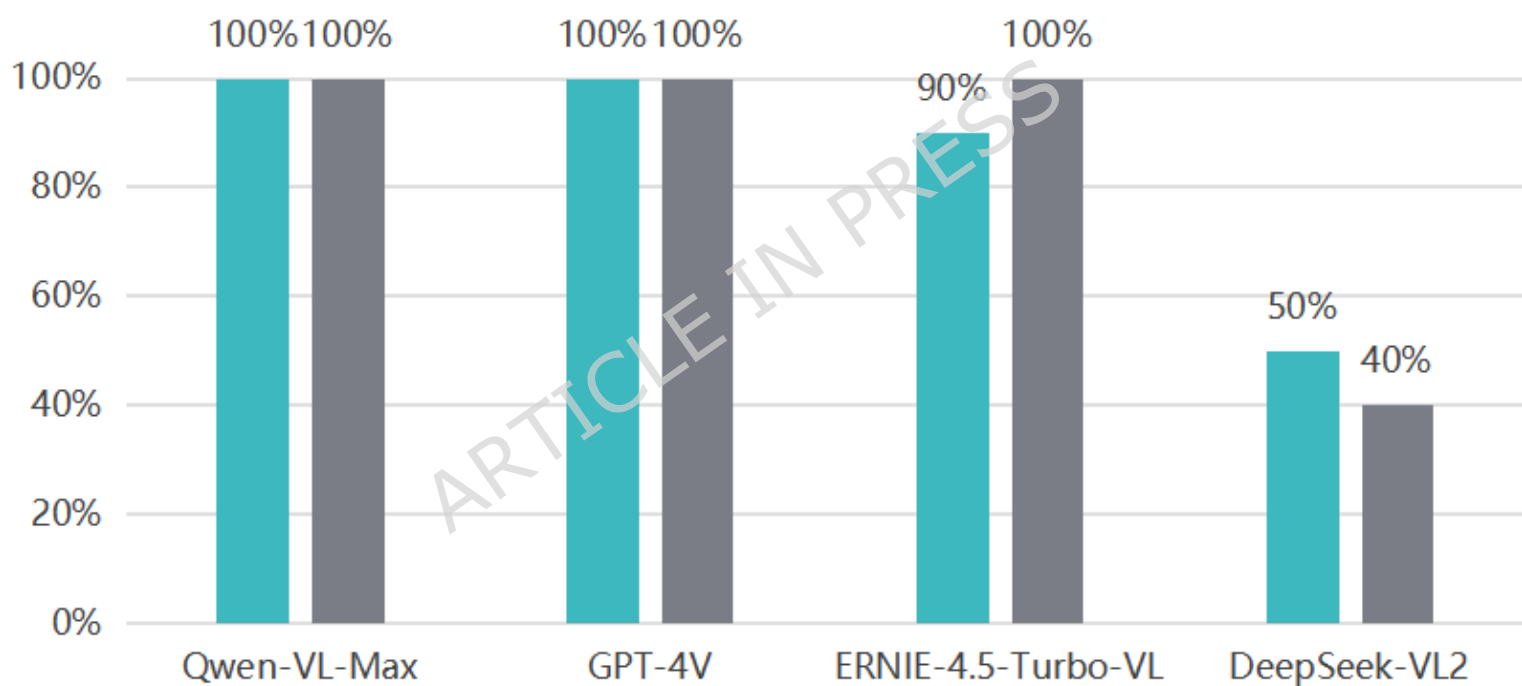
Visual-Textual Alignment Test

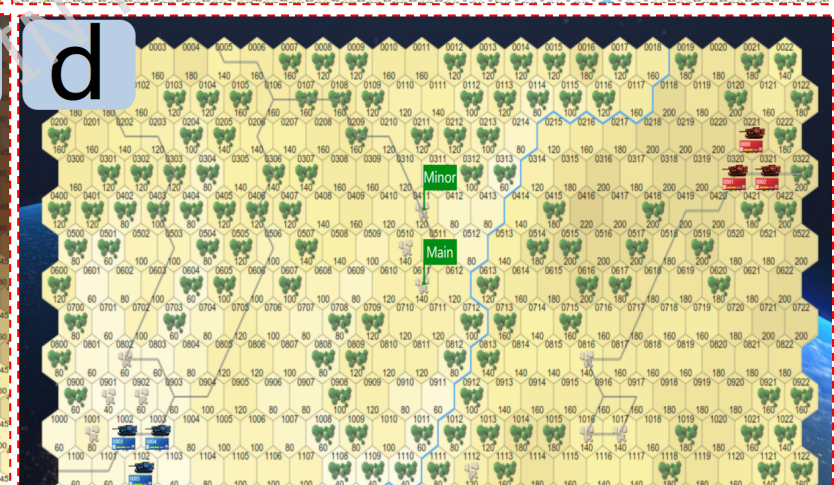
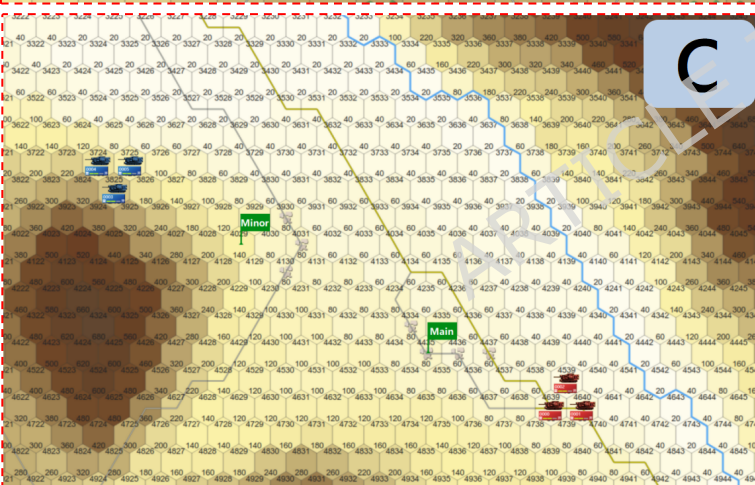
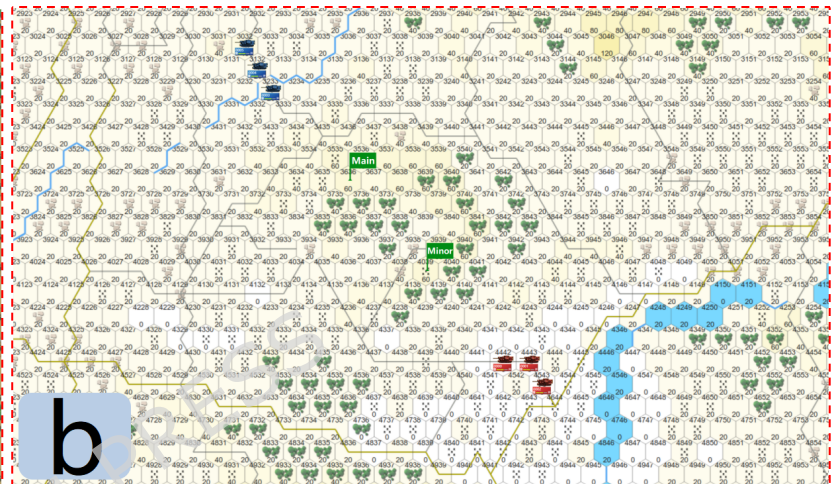
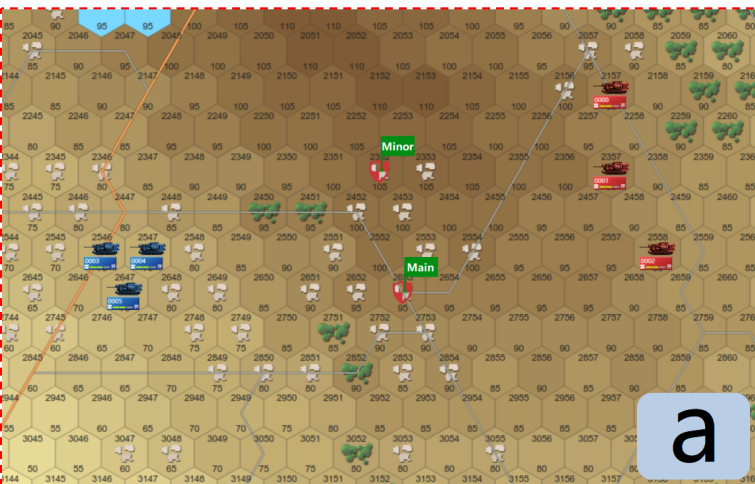
■ Geographical Element Identification Accuracy ■ Elevation Identification Accuracy



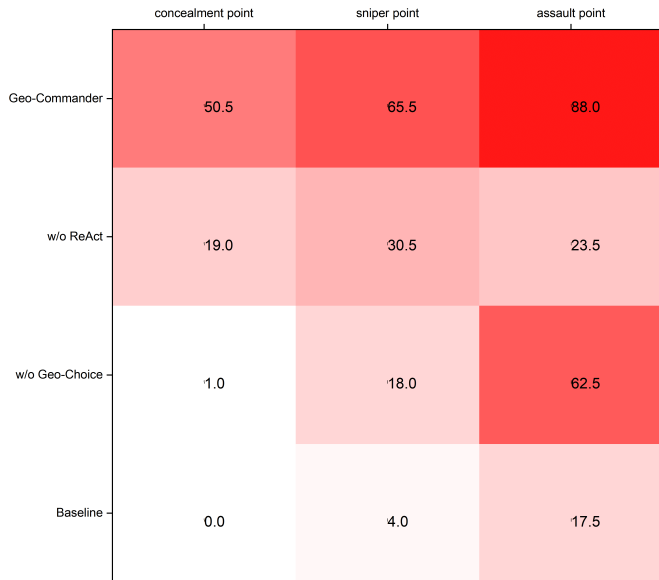
Misreading Elimination Test

■ Geographical Element Identification Accuracy ■ Elevation Identification Accuracy

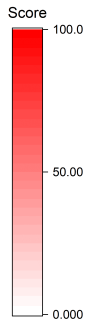
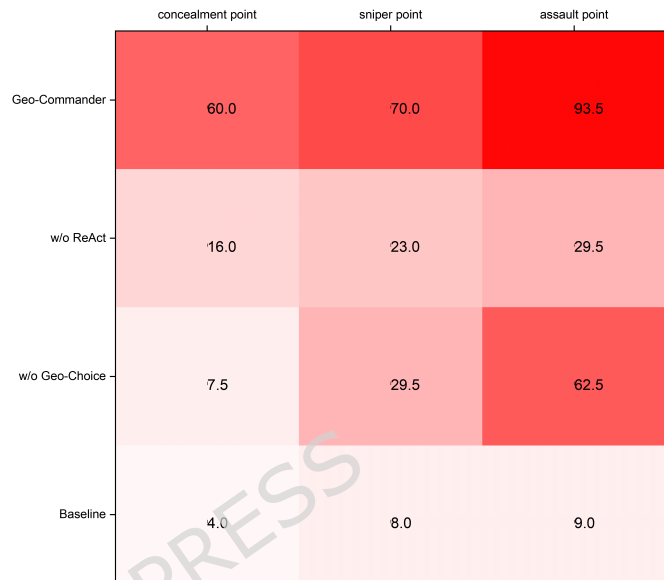




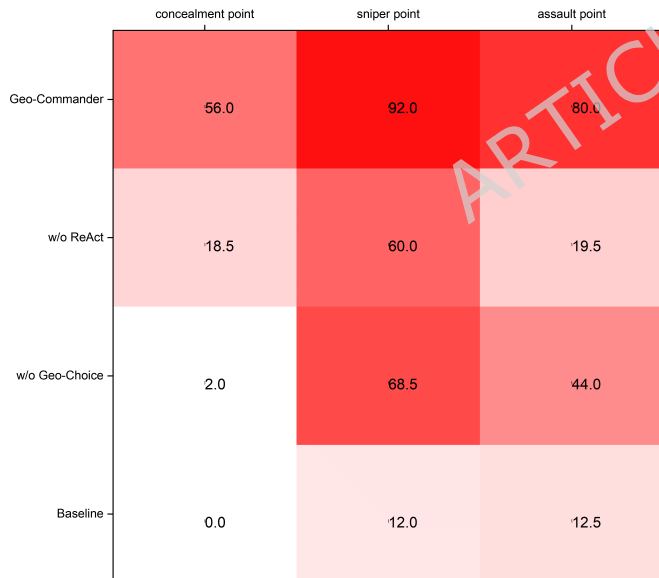
Qwen-VL-Max



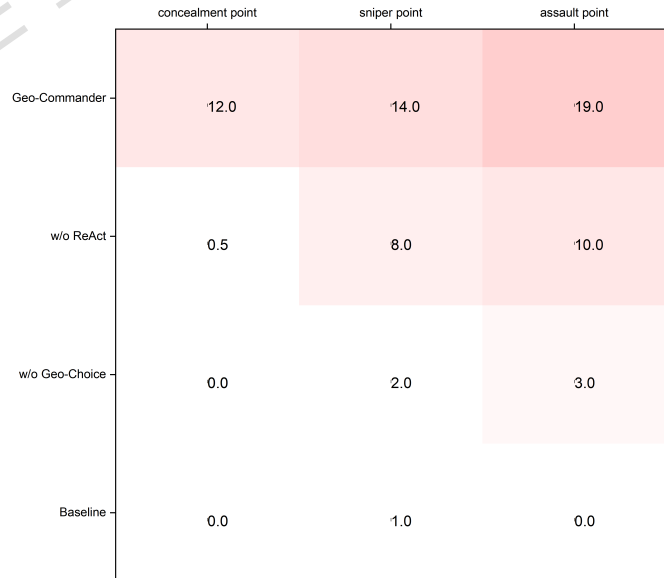
ERNIE-4.5-Turbo-VL

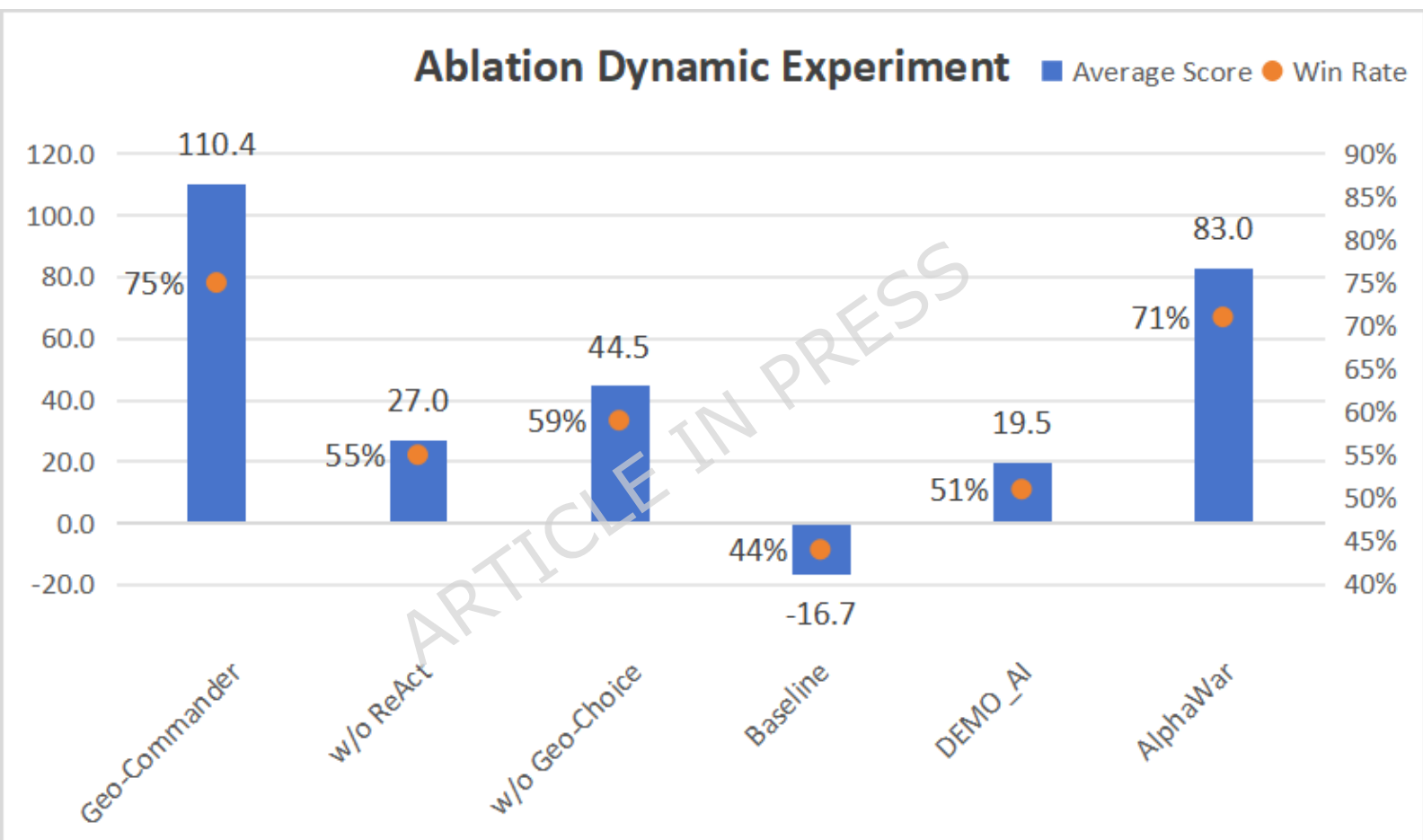


GPT-4V

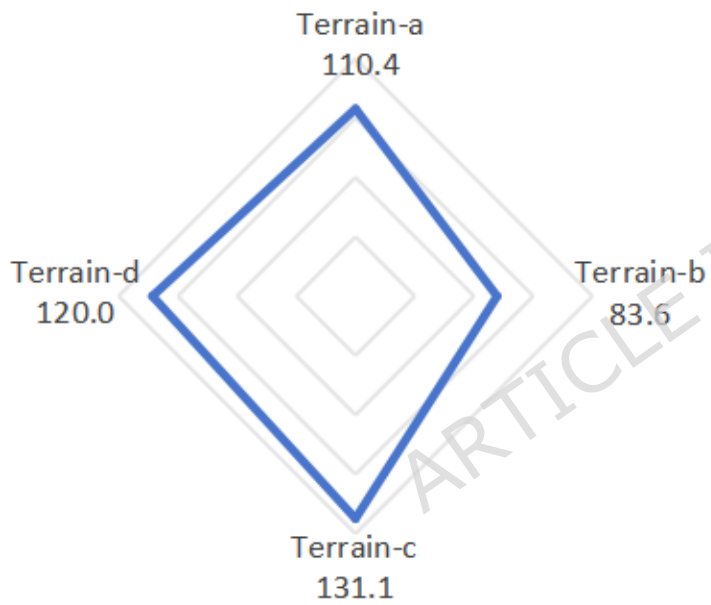


DeepSeek-VL2

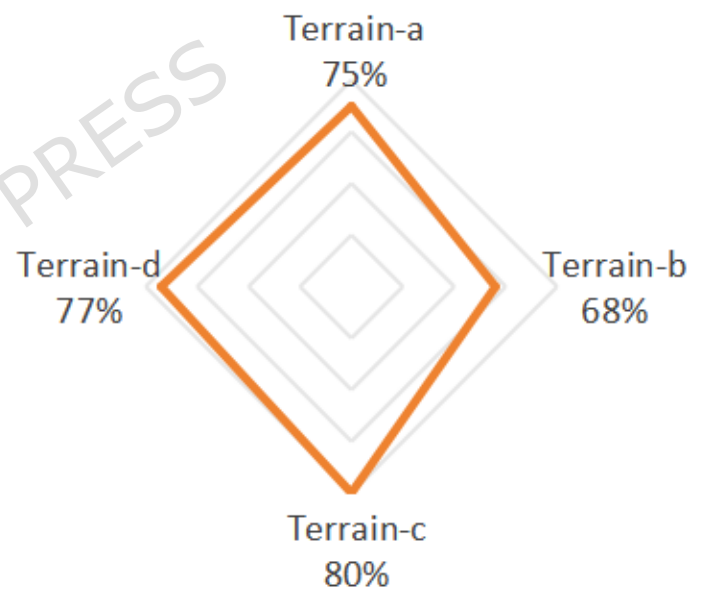




Generalization Dynamic Experiment Average Score



Generalization Dynamic Experiment Win Rate



Generalization Dynamic Experiment**Average Score**Terrain-a
128.9Terrain-d
108.4Terrain-b
134.3Terrain-c
119.5**Generalization Dynamic Experiment****Win Rate**Terrain-a
79%Terrain-d
74%Terrain-b
80%Terrain-c
77%

Table 1. Grid Point Quality Rating Table (Terrain-a)

	Optimal	Suboptimal
Concealment Point	(24,56); (25,55)	(21,54); (21,55); (21,56); (22,55); (23,54); (23,55); (24,55)
Sniper Point	(23,52); (25,54); (26,52); (28,52)	(26,53); (23,53); (24,52); (24,53)
Assault Point	(24,52); (25,54); (27,53)	(25,53); (26,52); (27,52)