Large Language Models in Military Planning and War-gaming: Project Pitch

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

Abstract

This research explores the potential of Large Language Models (LLMs) to enhance Course of Action (COA) analysis in military planning processes. Current COA analysis methods, whether analog or computer-simulated, are resource-intensive and limited by personnel experience levels. This study proposes a novel "Wargaming Tool" that leverages LLMs to automate both simulation creation and analysis. The tool incorporates an LLM interpreter for scenario input, a custom Gymnasium environment for simulation, an LLM-based leader for decision-making, and a text-based output analyzer. Using Anthropic's Claude LLM and employing prompt engineering techniques including XML tags, chain of thought, and multi-shot prompting, the tool focuses on a US Army Light Infantry Platoon conducting a movement to contact. Evaluation involves domain experts using modified US Army Task Evaluation and Outlines. This proof-of-concept aims to demonstrate LLMs' capacity to streamline military planning, potentially leading to more efficient and effective decision-making processes. The project's outcomes, including documented prompt engineering processes and simulation code, will be made available via a GitHub repository, serving as a foundation for future developments in AI-assisted wargaming and decision support systems.

CCS Concepts

• **Computing methodologies** → **Interest point and salient region detections**; **Simulation types and techniques**; **Neural networks**; • **Applied computing** → **Military**.

Keywords

War-Gaming, Course of Action (COA) Analysis, Large Language Model (LLM)

ACM Reference Format:
Stephen Moore. 2018. Large Language Models in Military Planning and War-gaming: Project Pitch. In *Proceedings of Intelligent Systems: Robots, Agents, and Humans (CAP-6671 '24)*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

CAP-6671 '24, August 19 – December 05, 2024, Orlando, FL

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

1 Project Description

As discussed in the introduction section of the project's literature review, COA analysis is a critical step within MDMP because it identifies vulnerabilities and critical decisions with the COAs. From the conceptual process outlined in US Army doctrine [5] [3], Army Officers and Staff Non-Commissioned Officers (NCO) employ one of the two mechanisms to conduct COA analysis: either some form of analog analysis around a map or terrain model or using a computer simulation like the Division Exercise Training and Review System (DXTRS). The analog analysis can be thought of as playing a board game, where human players on opposing sides conduct turn-based actions that are adjudicated by some form of manual predetermined rule-based procedure. The computer simulated analysis can be thought of as playing a computer game, where the human controlled actions still exist – but the adjudication process is automated. Both mechanisms are resources intensive, consuming both time and personnel, and are limited by the experience level of the personnel involved.

This project aims to demonstrate the potential in leveraging LLMs to not only automate the creation of a simulation, but to also automate the analysis process with a "Wargaming Tool". The proof-of-concept tool created by this project will operate in the following manner (see figure 1 – Wargaming Tool Diagram):

- User Input: The process begins with the Officer or NCO providing a high-level description of the scenario to an LLM. This will include the type of units involved (both friendly and enemy), the type of operation being conducted, and the area of operations (environment).
- LLM Interpreter: This component takes the user's input and converts it into a format that can be used to initialize the simulation.
- Middleware: This is the bridge between the LLM components (LLM Interpreter and LLM Platoon Leader) and simulation. It initializes the simulation, provides information about the current state of the simulation to the LLM Platoon Leader, and implements the actions directed by the LLM Platoon Leader.
- Simulation: This component is a custom Gymnasium environment that simulates actions – counteractions of the COA. The Gymnasium environment allows for customized terrain, unit configurations, and tactics. It feeds specifically formatted information about the change of the state to the middleware component.

- LLM Platoon Leader: This component represents the LLM that acts as a platoon leader, which makes tactical macro-level decisions based on the current game state.

- The cycle of state update, decision making, and action execution continues until the simulation ends.

- Text-Based Output Analysis: After the simulation runs, this component analyzes the outcomes and presents them in a human-readable format.

2 Project Plan

The primary goal for this project is for other Officers or Staff NCOs to build upon this foundation for their specific unit's needs. Most of these units do not have access to extensive computing resources to train their own LLM; therefore, approach to develop this COA analysis tool is through the prompt engineering an existing LLM. The LLM selected for this project is either Anthropic's Claude Sonnet or Opus due the capability of the LLM and the existence of paid subscription.

The three strategies that this project will employ are: XML tags, chain of thought, and multi-shot prompting.

- The XML tag method will be used to help interpret the simulation initialization parameters. Various maps, unit compositions, and tactics policies (either rule-based or MARL-based) are a few examples of the parameters that use this method due to need to reduce the errors that could be caused by the LLM misinterpreting critical parts of the prompt. Additionally, it provides the flexibility to easily update any of the parameters as tactics or environments change.

- The chain of thought method will be used to help develop macro level actions for the platoon leader component. This method is designed to break down complex analytical tasks into a structured step-by-step process. This is ideal for the platoon leader component because the Army has multi-stepped decision-making processes that govern macro-level actions.

- The multi-shot prompting method will be used to provide the output analysis from the simulation. This method is ideal for tasks that require adherence to structured formats.

Given the complexity of military operations and the timeline constraints associated with this project, the echelon and task associated with this tool will be limited to a standard US Army Light Infantry Platoon conducting the task of movement to contact. Future iterations of this tool should include a wider array of tactical tasks and unit configurations to account for the vast array of tactics that can be employed.

To evaluate the tool's effectiveness, a small sampling of domain experts from the University of Central Florida will conduct a heuristic evaluation using modifications of three US Army Task Evaluation and Outlines (TE&O). These TE&Os are used during live training exercises to evaluate a units' or leaders' adherence to a task's doctrinal standards. A combination of the drills "react to direct fire", "conduct hasty attack", and "break contact" will be used [4].

3 Additional Tools

In addition to the resources discussed in the literature review, there are two resources that will help in the generation of the code for this project:

- Support documentation provided by Anthropic on how to prompt engineer their LLM.[1]
- Google Colab Notebook provided by Anthropic on a tool for prompt generation. [2]

4 Team

The version of the tool that this project is designed to develop, is an individual based project. Future versions of the tool will require collaboration with multiple personnel and agencies to better design the environment and tactics employed.

References

[1] Anthropic. [n. d.]. *Prompt engineering overview.* https://www.anthropic.com/prompt-engineering Accessed: [06 October 2024].
[2] Anthropic. [n. d.]. Prompt Generator. https://colab.research.google.com/drive/1MIuZ5pjKDWjY3ok20YtVyRwXPfKNOEgj Google Colab notebook, Accessed: [06 October 2024].
[3] Center for Army Lessons Learned. 2020. *How to Master Wargaming: Commander and Staff Guide to Improving Course of Action Analysis.* CALL Publication 20-06. United States Army Combined Arms Center. https://usacac.army.mil/sites/default/files/publications/20-06-how-to-master-wargaming-public.pdf
[4] Headquarters, Department of the Army. [n. d.]. Army Training Network (ATN). https://atn.army.mil/. Accessed: [I06 October 2024].
[5] Headquarters, Department of the Army. 2022. *Planning and Orders Production.* https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN36775-FM_5-0-001-WEB-3.pdf

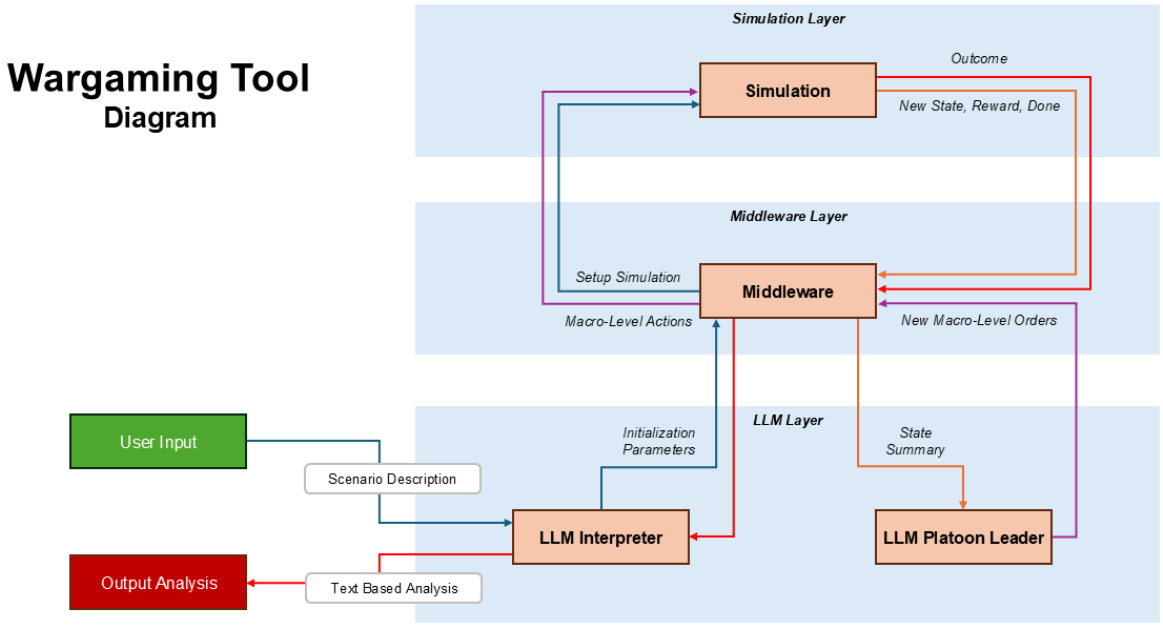

Figure 1: Wargaming Tool Diagram