Task Force Lima 101: CDAO Procurement Forum

Thursday, 30 November 2023
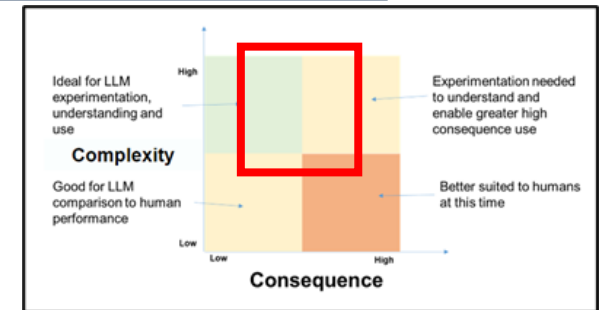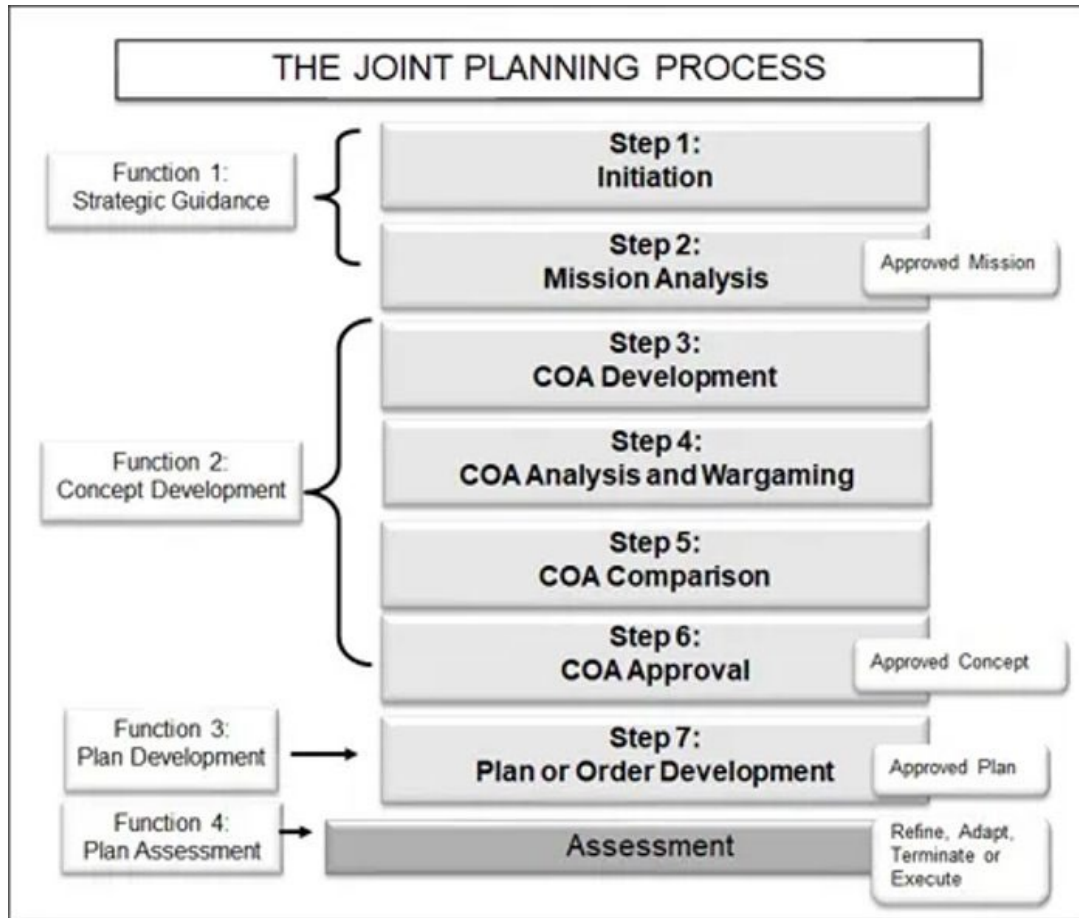
The overall classification of this presentation is **UNCLASSIFIED**

The Operational Planner's Current Workflow

User: LTC Bob Stackhouse is a plans officer at a CCMD leading a joint planning group (JPG).
Task: Collaborate with counterparts in the service components and other stakeholders to update an OPLAN.

THE JOINT PLANNING PROCESS

Function 1: Strategic Guidance
- Step 1: Initiation
- Step 2: Mission Analysis — Approved Mission

Function 2: Concept Development
- Step 3: COA Development
- Step 4: COA Analysis and Wargaming
- Step 5: COA Comparison
- Step 6: COA Approval — Approved Concept

Function 3: Plan Development
- Step 7: Plan or Order Development — Approved Plan

Function 4: Plan Assessment
- Assessment — Refine, Adapt, Terminate or Execute

- Workflow is cumbersome, text intensive, and requires significant time and coordination with numerous stakeholders
- Requires information from numerous data sources

Too much to do. Not enough time.

CDAO
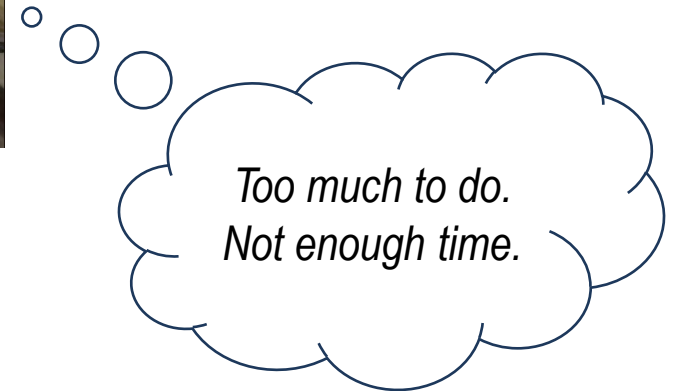
The Operational Planner's Workflow with an LLM

1

2

3

Products
1. **Strategic Estimate**
2. **Operational Approach**
3. **Commander's Estimate**

- **Problem**: The OPLAN, annexes, and appendices are immense documents that require a lot of time to manually review before updating

- **Data Sources**: The most current available information, including open source and classified data, to support targeting and fires deconfliction requirements.

- **Acute Issues**:
  - Fuel and ammo consumption
 - Synchronization across all JPG products
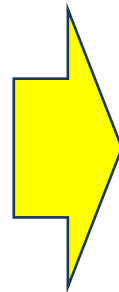
Prep Actions/LLM Inputs:

- JPG members scan analog artifacts and extract the textual content for use by an LLM.
- Data is imported from external authoritative sources to support IPB/IPE.

LLM Prompts:

- LogO uses iterative prompts to refine LLM outputs about fuel consumption based on the preferred COA
- LLM outputs offer citations to source docs
- Supports LogO planning for FARPs
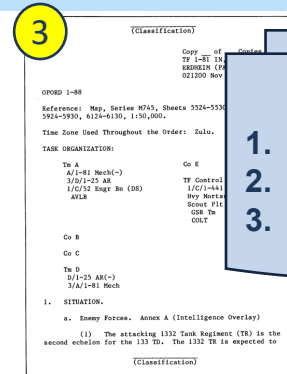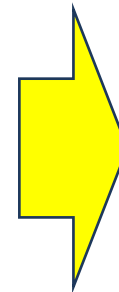
LLM Outputs: Text summaries to support initial drafts of products

Human-Machine Teaming: Planners closely scrutinize outputs to ensure accuracy/relevancy

Metrics: Speed/doctrinal accuracy of model; efficiency/effectiveness of JPG

Risks: Doctrinally inaccurate outputs; outputs generated using misinterpreted data; assumed model legitimacy for subject matter

Mitigations: Oversight by human subject matter experts; reinforcement learning via human feedback

CDAO

Problem Statement

DoD does not fully understand how best to maximize the opportunities and minimize the risks of generative AI, including large language models (LLMs), to enhance warfighting and enterprise requirements.

"This moment, which may not last, presents the Department of Defense (DoD) with a crucial opportunity to accelerate two of its most significant transformations – preparing for the future character of conflict and strengthening our military overmatch against our rivals, especially the People's Republic of China (PRC). The establishment of Task Force Lima by the Department of Defense on August 10, 2023, was a critical recognition of the importance of this moment…"

-- Generative AI: The Future of Innovation Power
Special Competitive Studies Project Report
September 2023

Why do we need a Task Force for Generative AI?

Christiane Amanpour interview with Dr. Craig Martell, CDAO,
7 August 2023

- "From a <u>science perspective, I like where we are. From a product perspective, I don't think we're ready</u>."

- "I'm very excited about artificial intelligence – as a scientist. Just not as an <u>end user</u> yet. I think there's a big gap between the science and deployment."

- "It's incumbent upon us to lay out a set of <u>use cases … and acceptability conditions</u>."

What is Task Force Lima?

TF Lima = DoD Generative AI Task Force

- Lima = **L**arge **L**anguage Model

- LLMs + other generative AI capabilities

Approved by Deputy Secretary of Defense on 10 Aug 2023 to…

- Accelerate promising generative AI initiatives and joint solutions…"

- Federate disparate developmental and research efforts into a DoD community of practice…"

- Drive education and build a culture of responsible implementation…"

- Ensure coordinated DoD engagement with interagency, international, educational, civil society, and industry partners…"

To advance a unified approach to responsible adoption and use of this technology, Task Force Lima will provide the framework and initiative for DoD to:

- Accelerate promising generative AI initiatives and joint solutions;

- Federate disparate developmental and research efforts into a DoD community of practice to accelerate innovation and implementation;

- Evaluate solutions across Doctrine, Organization, Training, Materiel, Leadership, Personnel, Facilities, and Policy;

- Drive education and build a culture of responsible implementation and use; and,

- Ensure coordinated DoD engagement with interagency, international, educational, civil society, and industry partners regarding responsible development and use of generative AI.

Task Force Lima will build strong partnerships across OSD and DoD Components, ensuring we collectively learn, responsibly adopt, and diligently advance the security and safety of generative AI. I ask all Components to partner with the Task Force and empower their RAI leads to incorporate its lessons. This technology has the potential to be transformative. We must leverage it to advance our Nation's security, in a responsible manner consistent with our values.
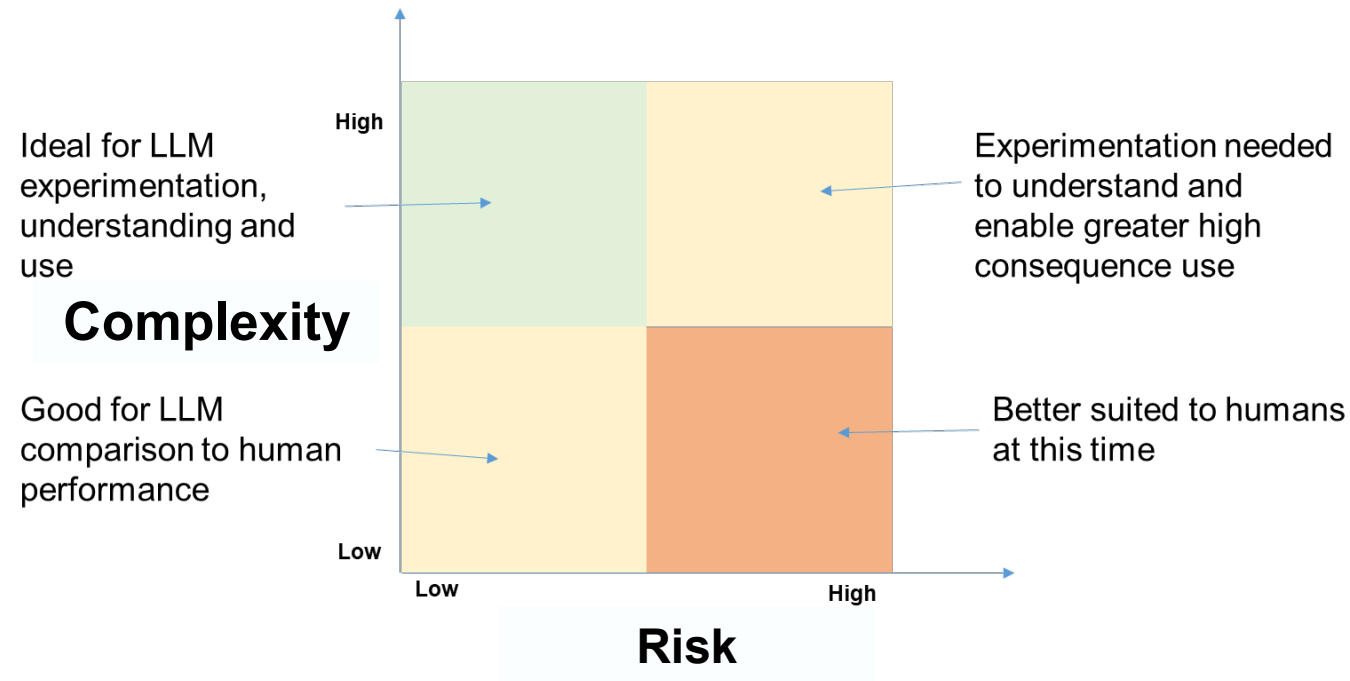
'A' Way to Conceptualize Gen AI/LLMs

Understanding the <u>users'</u> <u>experiences</u> is critical to developing the right capability, informed by data-driven empirical evidence.

Users must be aligned to the <u>most impactful, most realistic</u> use cases for generative AI.

Generative AI is <u>not the solution to solve every problem</u>.

CDAO will work with DoD components to refine <u>priority use cases</u>, develop <u>metrics</u> to measure <u>mission effectiveness</u>, and align expertise and support accordingly.

Ideal for LLM experimentation, understanding and use

Good for LLM comparison to human performance

Complexity

High

Low

Experimentation needed to understand and enable greater high consequence use

Better suited to humans at this time

Low   High

Risk

CDAO

Lines of Effort

LOE1: Learn

- Info Collection, Analysis, Management and Dissemination
- Evolve Communities of Interest into Communities of Practice
- Use Case Repository, Lessons Learned Library, Collaborative Environment
- Influence Training and Education
- Engage Industry and Academia to Inform Future Requirements

LOE2: Accelerate

- Establish Partnerships and Support Experimentation
- Access to Tools, Infrastructure, Contracting Mechanisms, and Subject Matter Expertise
- Highlight Exemplar Use Cases to Support Similar Initiatives Across Enterprise
- Resourcing Recommendations for Promising Initiatives and High-Payoff Capabilities
- Acceptability Criteria

LOE3: Guide

- Develop Policies, Regulations, Toolkits, and Other Guidance Materials
- Interim Guidance for Use of Generative AI
- Evaluation Framework/Metrics
- Cybersecurity/AML; Data Quality
- Traceability and Understandability
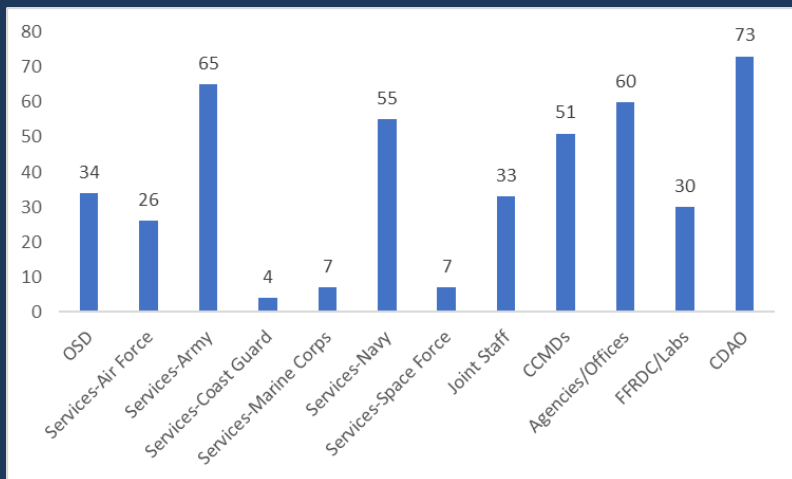- Human-Machine Teaming/Trust

LOE4: Messaging

Engagement and Partnerships

DoD Contacts
491 Total

TF Lima Working Group	**TF Lima Senior Council**
- DoD-wide (O-4/O-5 Level)	- DoD-wide (O-6+ Level)
- Meets Monthly	- Meets Quarterly

Engagement via TF Lima Battle Rhythm
- Senior Council (13 Sep)
- Working Group (11 Oct)
- Working Group (8 Nov)
- 184 Generative AI Use Cases Submitted

US Government
6 Total
White House, Dept of State, Virginia

Interagency Engagement Plan

International
4 Total
UK and Australia

International Engagement Plan

Industry
266 Total
122 Companies

Academia
27 Total
12 Institutions

- **CDAO Request for Info (6 Sep – 9 Oct)**
  - *103 White Papers submitted*
- **Industry/Academia Challenge Day (27 Sep)**
- **Procurement Forum Day (30 Nov)**

How Industry Can Help

Dr. Bill Streilein
Chief Technology Officer
CDAO

Call for Information: LLM Maturity Model Ideas

Open Call for Large Language Model Maturity Model Ideas

Introduction

The CDAO is seeking ideas related to the development of a Large Language Model Maturity Model (LLM MM) that could be used to assess the appropriateness of LLM solutions for given mission use cases and workflows. Selected LLM MM ideas will be the centerpiece of the LLM track at the upcoming Advantage DoD 2024 Symposium and will support community discussion aimed at improving LLM performance and suitability. The model should consider LLM capability in the context of classes of use cases, such as summarization, analysis, conversation, etc., and relate to LLM performance characteristics, such as accuracy, completeness, traceability, etc. The model should specify acceptability criteria for use of LLMs at different levels of capability for different use cases. For example, in the *notional* LLM Maturity Model depiction presented in Figure 1 below, Analysis use cases (5th column) may require a high-level (green cell) of accuracy and completeness, but a low-level (red cell) of interactivity.

- Industry feedback sought for maturity model to assess appropriateness of LLM solutions for DoD use cases/workflows

- Address classes of use cases, i.e., summarization, analysis, conversation

- Relate to performance characteristics, i.e., accuracy, completeness, and traceability/transparency

- Best responses will be showcased during Advantage DoD Symposium on 20-22 Feb 2024

- Deadline for submissions: **7 Dec 2023/2359 ET**

Tradewinds LLM Maturity Model Call for Information:
https://www.tradewindai.com/opportunities/cln26it8j0002mf083nyhw98r

Way Ahead

- 30 Nov: CDAO Procurement Forum (Wash, DC)

- 7 Dec/2359 ET: Industry LLM Maturity Model Submission Deadline

- NLT 31 Dec: Publish Initial Evaluation Framework

- Dec 2023-Mar 2024: LLM Experiments Across DoD

- 20-22 Feb 2024: Advantage DoD 2024 Symposium (Washington, DC)

DoD

Academia

Industry
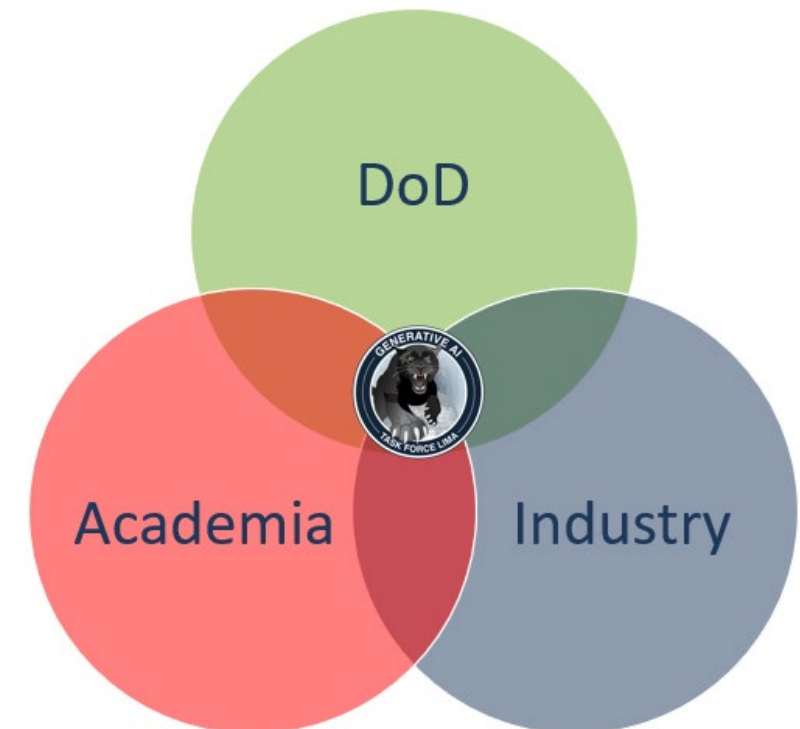
TF Lima Organizational Email:
osd.cdao.tf.lima@mail.mil

CDAO