## HR0011SB20254-07 Improving Battle Planning through AI Frequently Asked Questions (FAQs)

- Are there particular types of scenarios or environments that you're most interested in exploring during this effort?
   A: The government is interested in the ability to develop compositional models reflective of the underlying dynamic systems that can be composed for multiple scenarios.
- Could you share a bit about any specific challenges or pain points you've observed with current COA adjudication methods?
   A: The primary pain points are speed and accuracy for scaling purposes. Using reduced order models the government seeks to accelerate adjudication for horizontal and vertical scaling to many times faster than real time.
- Is there a preference or requirement regarding integration and composability with existing simulation or war-gaming frameworks?
   A: Responsive solutions will generalize for composition in future simulation and war-gaming frameworks over standard APIs.
- 4. The majority of the opportunity is tailored to specialized COA/ROM software. Will there be an opening for an architecture opportunity?
   A: Responsive solutions will be general and adaptable to many next-generation government COA generation and adjudication frameworks.
- 5. Could DARPA clarify whether the envisioned ROM adjudication architecture is expected to include adversarial modeling or co-evolution of blue/red COAs, or should the focus remain on evaluating blue COAs only?

A: The adjudication architecture should be capable of modeling both red, blue, and green COAs evolving at the same time in a ROM representation of a multi-physics simulation environment.

- 6. Does the government have a preference for how data heterogeneity is resolved within federated models (e.g., ontology alignment, schema harmonization, or data fusion heuristics), or is this left to the proposer?
  A: This is left to the proposer.
- 7. Are LLMs, knowledge graphs, or neural-symbolic hybrid systems considered responsive under the topic's intent, provided they are used to support composable ROM-based adjudication?

A: Provided the methods accelerate to many times beyond real time, and can be validated against trusted physics simulator, proposers may submit any technology solution they believe meets criteria. It should be noted, accuracy is a key metric, however.

- Would integration of reinforcement learning (e.g., for adversarial red COA generation or COA refinement under uncertainty) be aligned with the objectives of this topic?
   A: See Question 7.
- 9. Is there DARPA interest in incorporating models for psychological, civil, or information-based effects—such as morale, civilian sentiment, or influence operations—into the ROM adjudication process?
  A: Yes, DIMEFIL modeling are reponsive.
- 10. Can you confirm whether the government envisions real-time ROM-based adjudication in denied, degraded, intermittent, or low-bandwidth (DDIL) environments, or if centralized/cloud environments are assumed?
  A: The government envisions the ROM-based adjudication in next-generation war gaming environments.
- 11. Does DARPA expect direct interoperability with any existing planning systems (e.g., GCCS, JOPES), or is demonstrating API-driven composability sufficient?
  A: Demonstrating a generalizable API is more desirable than interoperability with a single planning system.
- 12. How will the accuracy of COA adjudication be measured during Phase II? Are baseline datasets or scoring references available?
  A: Against the original data sets or simulation environments used to generate the ROMs.
- 13. Are there specific expectations for modeling strategic effects across the DIMEFIL (Diplomatic, Information, Military, Economic, Financial, Intelligence, Law Enforcement) spectrum?

A: Expectations are primarily with respect to model fidelity.

- 14. Are there potential integration or transition pathways envisioned with programs such as SCEPTER, or expectations for working within classified, coalition, or interagency environments?
  A: Multiple transition pathways with DARPA and broader DoD needs are possible for a successful effort.
- 15. What echelon of command is envisioned as the primary user of the COA engine (e.g., tactical unit commander, joint task force, theater-level)? The COAs generated would differ significantly depending on strategic vs. tactical planning intent.
   A: Responsive proposals will address strategic, operational, and tactical levels.
- 16. Is the target battlespace primarily Army, Air Force, Navy—or joint/combined with coalition forces? We've observed significant differences when planning with ROK or other allied partners during SOF wargaming.

A: Responsive proposals will address joint, all-domain, warfare. Partners and allies modeling is also desired.

17. Will models explicitly account for variables such as weather, terrain, electromagnetic spectrum conditions, and time-of-day? These are key in real-world COA feasibility but often missing from abstract models.

A: Responsive proposals will address variables represented by the core modeling and sim environments being accelerated by ROM development.

18. How does the government envision modeling uncertainty or incomplete information? Is there an expectation to incorporate probabilistic fog-of-war mechanisms into ROM-based adjudication?

A: ROM-based adjudication systems should faithfully represent all characteristics of their original environments, including sensors, uncertainty, and incomplete information.

- 19. Will the government permit adaptive tolerance ranges in MOR algorithms that refine fidelity dynamically—without full retraining or manual recalibration—to enable COA adjudication at speeds five orders of magnitude faster than real time?
  A: Yes.
- 20. Can the government clarify whether there is a preferred modeling environment or simulation platform (e.g., AFSIM, WARSIM, JADC2 testbeds) with which the ROM architecture should interoperate in Phase II demonstrations?
  A: Proposed methods should be general enough to apply to any modeling and simulation environment.
- 21. To what extent should the proposed ROMs be able to ingest both real-time streaming inputs and pre-staged simulation data? Is there a performance or fidelity tradeoff that should be optimized for?

A: The proposed ROMs should be able to function as an accelerated version of the original environment for adjudication purposes, operating faster than real time, and with bounded error.

22. Does DARPA envision the composition operator and ROMs being reusable across operational domains (air, land, cyber), or is tailoring expected per domain due to differing physics models and planning logics?

A: The ROMs should be compositional across operational domains.

- 23. Are there specific standards or protocols DARPA recommends for model composability or federated model integration (e.g., HLA, FMI, DDS, or other simulation standards)?A: Proposers should ensure the generality of their APIs for interfacing.
- 24. For DIMEFIL representation in federated planning, is there any weighting or prioritization across the dimensions that the adjudication process should emphasize (e.g., M vs. I vs. F)?

A: Responsive proposals will validate their solutions across multiple aspects of the DIMEFIL to show generalizability.

- 25. Will the government provide sample COA datasets or adversarial simulation logs to calibrate initial ROM performance during early prototyping?A: If requested, COA datasets can be provided for existing M&S environments.
- 26. Is it acceptable for the Phase II prototype to rely on CPU-only inference pipelines for portability and containerization, or are GPU-accelerated ROM pipelines encouraged where latency is critical?

A: Provided the necessary acceleration is achieved with minimal error, no hardware restrictions exist.

- 27. Is the primary focus of this topic developing a high-speed simulation surrogate via reduced order modeling, or does it also include integrated battle management or decision-making components?A: The primary focus is on developing high speed surrogate modeling sufficient for
  - COA adjudication.
- 28. How are COAs expected to be represented and passed to the ROM as individual actions, full time-sequenced trajectories, or hierarchical policies?
  A: COAs will be represented using a well-documented hierarchical policy compatible with AFSIM and similar environments. Details on this representation will be released at kickoff, to contribute towards Month 3 steel thread goals.
- 29. What specific output is expected from the ROM for adjudication a scalar utility score, an end state, or a full state trajectory?
  A: A distribution characterizing likely state trajectories.
- 30. Should the ROM explicitly account for stochastic or adversarial (reactive) behaviors during evaluation?
  A: COAs will be provided for all sides, stochastic adversary behaviors are not anticipated for this effort.
- 31. What fidelity level is expected from the ROM, and will the government provide reference data or simulations for training/validation?
  A: Reference data and/or simulations will be provided. Proposals should be structured indicating expected fidelity levels achievable by their technologies, with evidence from prior work to justify claims.
- 32. Are there established standards or protocols (e.g., HLA, DDS, API specs) for integrating ROMs into federated planning architectures?
  A: A reference protocol will be provided at kickoff for integrating into the steel thread indicated by the Month 3 milestone.
- 33. Should we propose a Design Reference Mission (DRM) or is there on in mind?A: Proposal of a DRM is not required. The government intends to provide several example scenarios for evaluation.

34. What command echelon would this be expected to operate at - e.g. tactical, theater, strategic?

A: Responsive proposals will address strategic, operational, and tactical levels with a unified, generalizable, solution.

- 35. Can we assume COA generation would be an input to this effort, i.e. generated by other programs or efforts? Or do we need to generate the COAs as part of this effort?
  A: Multiple generated COAs for adjudication will be provided by the government for the scenarios in question. The government anticipates providing COAs on a scale of 100,000+.
- 36. For computational scale, how many different units or entities would be considered in the various courses of action?

A: Responsive proposals should propose methods capable of scaling to 100,000+ decision making entities in theater (roughly 10,000+ units per side).

37. What specific pain points in current COA planning/adjudication is this topic aiming to solve?

A: The primary aim of the SBIR topic is to accelerate existing adjudication methods to support real-time battle planning and war gaming needs from slower than real time, to orders of magnitude faster than real time, while constraining error versus the original analysis cell.

- 38. What is the envisioned operational tempo and environment for COA planning (e.g., real-time joint operations, rapid-response tactical edge)?
  A: The goal is to support multiple use cases requiring faster-than-real time adjudication.
- 39. Which stakeholders and end-users (e.g., Joint Planning Staff, AOC, Combatant Commands) will directly use this solution?
  A: The solution is envisioned as part of an ecosystem of emerging tools for use by all echelons.
- 40. Are there existing battle planning systems/platforms (e.g., JADC2, ATAC, CPCE) we must integrate or remain interoperable with?A: The government will provide a common API for integration at kickoff.
- 41. What performance metrics will define success (e.g., adjudication speed, accuracy, scalability)?
  A: The primary metrics are on adjudication speed, accuracy compared to the original analysis cell, and scalability in number of units.
- 42. How is "five orders of magnitude faster than real-time" practically defined and measured?

A: Normal analysis cells adjudicate slower than real time, i.e. a 1 hr COA can take 3+ hours to adjudicate. Responsive proposals will accelerate this to many times faster than real time, i.e. adjudicating a 3-week COA in 500 seconds would yield a speedup > 3,500x.

- 43. What are the key risks and dependencies (e.g., data availability, integration timelines, classified models)?
  A: Proposers should identify key risks in their proposed methods. Analysis cells and data are already available at the CUI and higher level, and access will be granted at kick off.
- 44. What data access and simulation environments will be provided or need to be sourced independently?A:The government will provide access to data at kickoff, and pathways to obtain

A: The government will provide access to data at kickoff, and pathways to obtain licenses to responsive simulation environments.

45. What are the target pathways for transition to operational programs of record (PoRs) or acquisition offices?

A: Multiple pathways have been identified.

- 46. Is a government test team or customer base already identified to evaluate the Phase II prototype?A: Yes.
- 47. What is the required system architecture for federation (e.g., centralized, distributed, containerized microservices)?A: Containerized microservices.
- 48. Are there existing APIs, data formats (e.g., JC3IEDM, MIL-STD-2525), or message buses (e.g., DDS, ROS, Kafka) the system must use?
  A: The system will be based on a Restful API, and protobuf/ZeroMQ APIs backed by JSON and SQL. The schemas will be provided at kickoff.
- 49. What are the expected data formats, types, and resolutions (e.g., terrain, units, EW signals) to support the ROM generation?
  A: This information will be provided at kickoff for the test cases. Responsive proposals should generalize to multiple assumptions.
- 50. What types of physics or agent-based models are being reduced (e.g., mobility models, sensor effectors, attrition models)?A: All of the above.
- 51. Which ROM generation techniques are appropriate (e.g., POD, DMD, autoencoders, Gaussian process emulators)?

A: Responsive proposals will suggest a ROM generation technique they believe is responsive to the indicated problem space and should offer proof of feasibility as part of their Phase 1 performance supporting a direct to phase 2 award.

- 52. How will we handle the diversity of model fidelities and heterogeneous data sources? A: Numerous techniques exist for multi-scale modeling. Responsive proposals will outline why their proposed methods generalize best to adjudication with reduced order models.
- 53. What constraints exist for real-time inference on these ROMs (compute, latency, bandwidth)?
  A: Responsive proposals should bound their solutions with realistic requirements for evaluation by the government review team.
- 54. What methods will be used to ensure model composability (e.g., interface abstraction, shared ontology, standardized APIs)?A: Standardized APIs and interface abstraction.
- 55. How will we reconcile differences in time steps, spatial resolutions, or semantics across federated models?
  A: Responsive proposals will suggest methods from multi-resolution modeling methods to solve these and other challenges in generalizability.
- 56. What synchronization, validation, and conflict resolution mechanisms are needed for federated execution?

A: Responsive proposals will suggest methods for tackling problems in federated execution.

- 57. What level of classification will the final solution operate at (e.g., SECRET, TOP SECRET, SAP)?
  A: Successfully developed solutions will be anticipated to run at multiple levels of security.
- 58. What cybersecurity compliance is required (e.g., RMF, STIG, zero trust)?A: Successfully developed solutions will later address cybersecurity compliance.
- 59. How do we maintain model/data integrity when ingesting from disparate or third-party sources?

A: Responsive proposals will suggest methods from multi-resolution modeling methods to solve these and other challenges in generalizability.

60. What simulation testbeds or operational scenarios will be provided for prototype evaluation?

A: This information will be provided at kickoff.

- 61. How will we benchmark our ROMs for accuracy, speed, and stability versus full-order models?A: Proposed solutions will be validated against the original trusted analysis cells.
- 62. What validation datasets and ground truth will be available (e.g., historical scenarios, RED/BLUE outcomes)?
  A: Multiple examples of COAs and adjudications by trusted analysis cells will be provided at kickoff.
- 63. What load or stress conditions must the ROM adjudicator handle (e.g., # of COAs, # of federated models, scenario complexity)?
  A: Full, all domain, warfare at strategic, operational, and tactical levels, scaling to realistic theaters.
- 64. Can we integrate deep learning or neurosymbolic approaches to augment ROM generation or improve model fidelity?A: Performers may propose any technology they believe address the core goals of this opportunity.
- 65. Should we incorporate human-in-the-loop or explainable AI components to support trusted decision-making?A: Such capabilities are desirable but not required.
- 66. To ensure operators trust AI-based adjudication, should we provide transparency, traceability, or audit trails?A: Such capabilities are desirable but not required.
- 67. Do you expect mechanisms to be in place to allow fallback or override by human planners?A: Yes.