

DPA26BZ04-DV016

Fusion of Abstract Learning and Context-Optimized Neural-methods (FALCON) Frequently Asked Questions (FAQs) as of 10 Aug 26

- 1. Direct to Phase II readiness evidence:** For a responsive Direct to Phase II proposal, what level of prior “Phase I-type” evidence is expected to demonstrate readiness: peer-reviewed publications, internal technical reports with data and analysis, working software prototypes, benchmark studies on structured datasets, or customer/user demonstrations?

A: The listed examples of evidence are all valid. Be sure to explain why you think they meet Phase I criteria and how they support your Direct to Phase II proposal and justifications.
- 2. Definition of emerging Machine Learning methods:** How broadly should “emerging Machine Learning methods for structured data” be interpreted? For example, would neuro-symbolic learning, probabilistic graphical models, causal inference, tabular foundation models, decision-theoretic models, graph neural networks, or backpropagation-free learning methods all be considered within scope if they are compared against the current state of the art?

A: We originally considered structured data as anything in a tabular format (i.e. databases, spreadsheets). Graphs, networks, and relational databases could likely be relevant as well. Be sure to clearly define the structured data you intend to utilize in your proposal and how it can help us achieve our objectives. Consider methods that address or enhance how machine learning algorithms process, reason with, and learn from these structured formats. The methods you list (among others) could be candidates.
- 3. Mission and dataset priority:** Should offerors prioritize enterprise analytics, battlefield analytics, engineering/design data, manufacturing data, economic/social/geographic data, or a general architecture that can be evaluated across multiple structured-data domains?

A: The goal of this effort is to combine advanced machine learning (ML) methods that can be computationally efficient in structured data with large language models (LLM) that are general and can extract context from data. The aim is to derive powerful and efficient technology for interactive statistical analysis of large-scale data for many general use-cases. Please justify why you picked the missions and datasets you do based on this aim, and how they will support development of a general architecture.
- 4. Structured and unstructured data assumptions:** What data characteristics should be assumed for evaluation, including dataset scale, number of tables, schema complexity, missing-data rates, temporal structure, text length, label availability, data sensitivity, and whether classified or controlled unclassified information will be involved after award?

A: Neither classified nor CUI information will be involved after the award. Please feel free to introduce and justify the data assumptions included in your proposal. Please explain how these assumptions will support program aims.
- 5. Large Language Model boundary conditions:** Does the Government prefer open-source Large Language Models only, or would proprietary/commercial models be acceptable for early feasibility if the Phase II architecture can later transition to an on-premises, air-gapped, or Government-approved deployment?

A: Open-source LLMs or proprietary/commercial models are acceptable. Please explain how your choices support program aims.

- 6. Evaluation metrics and baselines:** What metrics should be treated as most important for proposal evaluation: predictive accuracy, insight discovery, statistical validity, computational efficiency, latency, memory footprint, generalization across datasets, reproducibility, analyst workload reduction, or improvement over Machine Learning-only and Large Language Model-only baselines?

A: The evaluation of the combined method should demonstrate accuracy relative to ground truth and document the improvement over SOTA methods (ML only, and LLM only) as well as demonstrating generality. Please explain why you believe you can accomplish these metrics, as well as any others, you include in your proposal.

- 7. Hallucination mitigation and reproducible analytic traces:** What level of evidence is expected to demonstrate hallucination mitigation and verifiable analytic traces: source-linked reasoning, executable notebooks, audit logs, statistical provenance, confidence scoring, formal uncertainty quantification, human approval checkpoints, or repeatable end-to-end workflow replay?

A: Please explain your proposed hallucination mitigations and proposed evidence for reproducible analytic traces.

- 8. Interactive analysis workflow:** Who should be treated as the primary user for the interactive analysis workflow: data scientist, mission analyst, engineer, program manager, intelligence analyst, or non-technical decision maker; and what actions should the user be able to take when incorporating new insights during analysis?

A: Please feel free to introduce and justify primary users included in your proposal. Please explain how these designs will support program aims. Please justify why you picked the users you do based on this aim, and how they will support development of a general architecture.

- 9. Enterprise-scale deployment constraints:** What deployment constraints should be assumed for Phase II, including cloud versus on-premises operation, disconnected or air-gapped environments, graphics processing unit availability, central processing unit-only operation, security accreditation, data-ingestion interfaces, and integration with existing Government data platforms?

A: Please focus on developing a more advanced, more general, and more technical solution to the program. Focus on doing more faster.

- 10. Phase II demonstration and transition expectations:** For the required Phase II demonstrations using at least two datasets from different areas, should offerors propose their own representative datasets, expect Government-provided datasets, or plan for a hybrid approach; and what would constitute a compelling Month 6 initial demonstration, Month 11 enterprise-scale demonstration, and final software delivery?

A: Yes, please propose your own representative dataset/ Propose demonstrations that will encourage the big players in industry and USG to want to engage with FALCON.

11. Should the LLM primarily explain ML outputs, or is it also expected to participate in inference?

A: The goal of this effort is to combine advanced machine learning (ML) methods that can be computationally efficient in structured data with large language models (LLM) that are general and can extract context from data. The aim is to derive powerful and efficient technology for interactive statistical analysis of large-scale data for many general use-cases. Please justify why you picked the approach based on this aim, and how your approach will support development of a general architecture.

12. What operational or analytical workflow are you ultimately trying to improve by combining ML and LLMs? Is there a specific Phase I use case you would like proposers to focus on?

A: Please feel free to introduce and justify primary users included in your proposal. Please explain how these designs will support program aims. Please justify why you pick the users you do based on this aim, and how they support development of a general architecture.

13. The solicitation states, "Evaluate the combined architecture and methods in datasets drawn from multiple applications. This may focus on tabular data in enterprise or engineering applications." Could you elaborate on the expected scope of these datasets?

A: We originally considered structured data as anything in a tabular format (i.e. databases, spreadsheets). Graphs, networks, and relational databases could likely be relevant as well. Be sure to clearly define the structured data you intend to utilize in your proposal and how it can help us achieve our objectives.

14. What do you mean by "verifiable and reproducible analytic traces"?

Are you referring to provenance tracking, full audit trails, explainable reasoning, or something else?

A: Please explain how your solution approaches verifiable and reproducible analytic traces and how it supports program aims. The examples you provided (among others) may be relevant.

15. The solicitation indicates a preference for open-source models. Are there particular open-source LLMs or deployment constraints we should consider?

A: No. Open-source LLMs or proprietary/commercial models are acceptable, though open-source models are preferred. Please explain how your choices support program aims.

16. The solicitation mentions "interactive statistical analysis." What level of interaction is envisioned (e.g., real-time interaction versus post-analysis review)?

A: The goal of this effort is to combine advanced machine learning (ML) methods that can be computationally efficient in structured data with large language models (LLM) that are general and can extract context from data. The aim is to derive powerful and efficient technology for interactive statistical analysis of large-scale data for many general use-cases. Please justify why you picked the level of interaction for your solution.

17. Should the proposed solution target cloud, on-premises, or both deployment environments?

A: Please focus on developing a more advanced, more general, and more technical solution for the program. Focus on doing more, faster.

18. In this context, what would constitute an LLM hallucination?

A: Please define LLM hallucination risks and mitigations based on your proposed solution.

19. Will representative or government datasets be made available for development or evaluation?

A: Please propose your own representative datasets.

20. Does a proposal centered on the fusion architecture, built around an existing emerging ML technique instead of a newly developed model, meet the topic objectives?

A: Developing a new ML model or method as part of Phase II, and selecting, adapting, and integrating existing emerging ML methods with one or more LLMs are acceptable. The aim is to derive powerful and efficient technology for interactive statistical analysis of large-scale data for many general use-cases. Please justify why you picked the approach based on the program aim, and how your approach will support development of a general architecture.

21. We believe a combination of LLMs and efficient, low-power ML would be the most useful approach for our specific developments and existing client base. Would our approach be of interest?

A: Potentially. The goal of this effort is to combine advanced machine learning (ML) methods that can be computationally efficient in structured data with large language models (LLM) that are general and can extract context from data. The aim is to derive powerful and efficient technology for interactive statistical analysis of large-scale data for many general use-cases.

Please note the emphasis above on “many general use cases”. Proposers will need to identify one or more promising ML methods and develop the software for combining it with one or more available LLMs. Open-source and proprietary are acceptable, but open-source is preferred. The final demonstrations are to be done with at least two datasets from different areas. We are interested in solutions and demonstrations that will encourage the big players in industry and USG to want to engage with FALCON.

22. For the Direct-to-Phase-II (DP2) feasibility documentation, can you confirm whether "the research team's prior comprehensive research experience" needs to be in a documented form? Our team has worked on several public sector projects in a consulting mode. But these are generally work performed for other government organizations.

A: All cited experience needs to be auditable.

23. Section 1.7 of the DoW BAA states feasibility documentation "cannot be based solely upon or logically extend from any prior or ongoing federally funded SBIR or STTR work." Can you clarify this please?

A: The intention of DP2 topics is to solicit feasibility work performed outside of the SBIR/STTR programs; if the work was performed on a prior SBIR, it can be extended via a sequential or initial Phase II based on the prior work vs. an entirely new topic.

24. If the proposing team intends to license (rather than own) IP underlying the ML methods described in the feasibility documentation, what level of documentation of licensure is required at proposal submission versus at award negotiation?

A: Please describe the planned licensure process in the proposal.

25. Is the \$1,000,000 base award amount / \$500,000 option amount listed for DV016 a hard ceiling?

A: yes

26. If the proposing team plans to use open-source LLMs as described in the Phase II objective, is any documentation of licensing terms/compliance for those open-source models expected in Volume 5 (Supporting Documents)?

A: Please justify why you picked the models based on the program aim, and how this approach will support development of a general architecture. Please describe how your plans to license your solution complies with the open-source models licensing and compliance terms.

27. The topic specifies prior research "should have been conducted in the last three years." Is this window measured from the proposal submission date or from the topic's pre-release date (July 1, 2026)?

A: The proposal submission date will work.

28. What level of evidence is expected to demonstrate "advantages relative to SOTA ML methods" — is a benchmark comparison against named SOTA methods required, or is a qualitative/literature-based comparison sufficient at the feasibility stage?

A: Either level of evidence could work. Please explain how the level of evidence you provide make you ready to accomplish the program aims. Reports which provide data, clearly present the analysis done, and provide evidence of scholarly impact will be strongly preferred.

29. Does "reports which provide data, clearly present the analysis done, and provide evidence of scholarly impact" require peer-reviewed publication, or would internal technical reports, preprints, or benchmark submissions (e.g., NeurIPS workshop papers) be acceptable evidence of scholarly impact?

A: Evidence of scholarly impact measures the influence and reach of your academic research. Please feel free to provide internal technical reports, preprints, or benchmark submissions and provide evidence on how they had scholarly impact. Reports which provide data, clearly present the analysis done, and provide evidence of scholarly impact will be strongly preferred.

30. Since the effort spans both structured/tabular ML methods and LLM integration, must the team's prior research history demonstrate experience in both areas, or is separate expertise in each area (e.g., one PI with ML-methods background, another with LLM background) acceptable for establishing feasibility?

A: Yes, cumulatively the team's research history should demonstrate experience in both areas.

31. What types of structure and unstructured datasets best represent the intended enterprise or battlefield analysis use cases?

A: The goal of this effort is to combine advanced machine learning (ML) methods that can be computationally efficient in structured data with large language models (LLM) that are general and can extract context from data. The aim is to derive powerful and efficient technology for interactive statistical analysis of large-scale data seen such as in enterprise or battlefield. We originally considered structured data as anything in a tabular format (i.e. databases, spreadsheets). Graphs, networks, and relational databases could likely be relevant as well. Be sure to clearly define the structured data you intend to utilize in your proposal and how it can help us achieve our objectives.

On the use case, please note developing solutions that can generalize across datasets from different domains is a priority for this effort. The final demonstrations are to be done with at least two datasets from different areas.

- 32.** Which performance dimension is most important: accuracy, new insight generation, computational efficiency, or interactive usability?

A: Metrics should encompass accuracy, new insight generation, computational efficiency, or interactive usability. Be sure to clearly define the metrics you intend to utilize in your proposal and how it can help us achieve our objectives. If you need to prioritize the metrics based on your solution, please explain how you prioritize them and why, and how this prioritization can help us achieve our objectives.

- 33.** Should LLM integration prioritize open-source models, government-controlled deployment, or model-agnostic architecture?

A: Open-source and proprietary are acceptable, but open-source is preferred. Developing solutions that can generalize across datasets from different domains is a priority for this effort. The final demonstrations are to be done with at least two datasets from different areas.

- 34.** May a proposer demonstrate sufficient feasibility through documented technical and operational results even when the work has not resulted in peer-reviewed academic publications, or does DARPA expect evidence of scholarly impact for a responsive proposal?

A: Yes. Please explain how the listed kind of evidence makes you ready to accomplish your proposed solution and the program aims. Reports which provide data, clearly present the analysis done, and provide evidence of scholarly impact will be strongly preferred. Focusing on data-driven evidence and associated analysis is preferred if you don't have evidence of scholarly impact. If you don't have evidence of scholarly impact, please feel free to explain how you've advanced Topic related science to date in other ways.