

GREEN Q&As HAVE BEEN PUBLISHED. BLUE HAS NOT

DARPA Mapping Machine Learning to Physics (ML2P) Program - Q&A

This document summarizes questions and answers related to the DARPA ML2P program, based on the DARPA-PS-25-32 solicitation posted to SAM.gov on September 23, 2025.

SCOPE AND ALIGNMENT WITH PROGRAM OBJECTIVE

- 1. How would you react to an abstract that does not address the full scope of the call, but focuses on what could be a component in a larger project?**

A: In accordance with Attachment A1, proposers should include all information in the template to constitute a fully conforming abstract submission. Abstracts that fail to address the content requirements of Attachment A1 may be found less favorable (weaknesses, significant weaknesses) during the evaluation process described in 6.3 of the solicitation.

- 2. Assuming sufficient proof of concept, is the following out of scope for ML2P: Developing new AI/ML technology that natively runs on very low power devices so that technology can also perform regression, classification, clustering, etc.; and can be benchmarked/mapped in.**

A: The ML2P program aims to improve power consumption and performance of machine learning (ML) models by preserving local energy semantics and tuning energy-performance objective functions. ML2P will explore a set of multi-objective functions, balancing the trade-off of joules with performance, covering a range of common ML tasks. Technical solutions should design experiments and collect the energy semantics of machine learning (ES-ML) to discover interactions between upstream optimizations and their downstream effects. New AI/ML technology other than hardware is welcomed. Note that the ML2P program objective is producing code, algorithms and documentation; so new hardware is not in scope.

- 3. Is developing novel low-power ML models in scope, or is ML2P singly focused on developing a rigorous power benchmarking/mapping technology for existing ML**

A: ML2P is focused on the entire model pipeline: not just the model itself but data ingest and preparation, model building and evaluation, and inference. In particular, ML2P is not concerned with the power used by a single component, but rather the power used as a result of interactions between each step or component of the process.

- 4. Is online learning in scope of the program or are only train-then-inference models of interest?**

A: Development teams may use any ML model or algorithm they choose, as long as they are able to capture the energy semantics of the entire ML pipeline (not just the energy used by the model).

5. Regarding discovery of new energy efficient algorithms, are distributed multi-node algorithms (like branch-train-merge mentioned at proposers' day) within scope?

A: Development teams may use any machine learning model or algorithm they choose, as long as they are able to capture the energy semantics of the entire ML pipeline (not just the energy used by the model).

6. My team was wondering if large language models were within scope for ML2P or whether other types of models (e.g., strictly classification or clustering models) were meant to be the focus?

A: Development teams may use any machine learning model or algorithm they choose, as long as they are able to capture the energy semantics of the entire ML pipeline (not just the energy used by the model).

7. We would like clarification on which software and hardware design/configuration variables fall within the scope of this program. Specifically, we are wondering whether different learning paradigms, such as distributed learning, federated learning, fine-tuning of pre-trained models, incremental learning, and prototype learning, are regarded as design variables subject to optimization. Given that these approaches exhibit distinct trade-offs between accuracy and energy consumption, it seems natural to consider and compare them within the ES-ML context. Furthermore, in the case of distributed or federated learning, should the energy costs associated with communication be explicitly included as part of the overall energy dissipation?

A: The ML2P program seeks to optimize energy usage across the entire ML lifecycle, making all design decisions within the ML pipeline in scope. This includes software, hardware, training paradigms, and activities with inherent tradeoffs within the ML pipeline. Specific learning paradigms like distributed, federated, fine-tuning, incremental, and prototype learning can be considered as design variables, considering the trade-offs between accuracy and energy consumption. In distributed or federated learning, the energy costs associated with network communication must be explicitly included in the overall energy dissipation, reflecting the program's focus on energy efficiency throughout the ML lifecycle and across optimization considerations.

8. Could the ML2P Team clarify whether the program permits heterogeneous compute pipelines, provided energy is measured consistently and results are comparable across platforms?

A: The solicitation emphasizes consistent energy measurement and comparable results across platforms; heterogeneous compute pipelines are permitted. The program's focus is on comparability through precise granular measurements in joules across hardware architectures, thus heterogeneous compute pipelines the energy measurements should be consistent and allow for meaningful comparisons.

- 9. With respect to domain-specific tasks, is a preference placed for including accuracy metrics that are standard in that domain alongside the program's specified task metrics? (capturing performance appropriately where it matters is the utmost priority for our proposal submission.)**

A: The program specifies A/P/R/F1 as performance metrics. While not explicitly preferred, including domain-specific accuracy metrics alongside the program's specified metrics is reasonable.

- 10. Please confirm that the program structure will be staged into two-stages: a 12 month with a go/no-go review at the end of the first period with an additional 12-month extension if successful.**

A: In accordance with the solicitation, the program structure is divided into two 12-month phases with a go/no-go decision at the end of Phase 1.

TECHNICAL REQUIREMENTS AND APPROACH

- 1. Can broad leeway be assumed when converting a trained model to specific hardware e.g., we can change the model architecture itself (or the model itself)?**

A: Technical solutions should focus on developing algorithms and software to convert trained or partially trained ML models for use on the different hardware types identified within the technical solution. The goal of ML2P is to create a way for ML model builders and users to choose from a list of ML pipeline options based on the energy resources they have available. This coupled with the key transition objective of ML2P – to make ML2P software the gold standard for ML construction and simulation of power usage and trade-offs – means that ML2P is seeking solutions that can be widely adopted and easily implemented. Any combination of model and physical hardware is within scope.

- 2. During the proposer's day, first order logic was mentioned as the formalism for ES-ML energy semantics representation. Are approaches not using logic to represent ES-ML considered out of scope?**

A: Approaches not using logic to represent ES-ML are not necessarily out of scope. The use of first-order predicates is strongly encouraged. ES-ML may be extensions of existing logic, calculus, or language (e.g., linear temporal logic, modal μ -calculus, systems modeling language). Technical solutions should develop a formal representation of energy semantics

for ML designed for machine readability; the use of first-order predicates is strongly encouraged.

3. **Can emulators be used as potential hardware choices? Built hardware may not exist for advanced compute fabrics, but emulation of individual operations exists. This may enable better hardware design using ML2P capability.**

A: The PS does not explicitly address the use of emulators as potential hardware choices. However, given the program's focus on accurate prediction of power and performance of future ML models and providing a foundation for simulation research in hardware design, the use of emulators could be a valid approach.

The hardware used for this program is expected to be physical hardware for the following reasons: a) The solicitation states that development performers should plan to deliver one unit of each hardware component used for processing and power measurement to the T&E Team for independent verification and validation; and b) The code and algorithms being produced by ML2P developers is intended to be widely adopted and used quickly, thus the hardware should already exist.

4. **Section 3.4 of the solicitations lists the key tasks in this program. In particular, Task 4 (Conversion Algorithms) calls for “Hardware agnostic Conversion” that would take (possibly partially) trained ML models and “convert” them for use on the hardware platforms that the selected teams design/propose. What is the expected output of this process? The actual software/code that is executed on the proposed hardware platform(s)? Or something else? Additionally, why should *each* team develop its own conversion algorithms? After all, if they are truly hardware agnostic, then a single one should suffice, especially in a low budget program? It would have a single set of conversion algorithms and be made available to all the teams. If the goal is to develop a mathematical framework of provably optimal ML that optimizes jointly for performance and energy as depicted in Fig 7, then, isn’t a complete conversion algorithm overkill? All that is needed is a mathematical function that takes the specification of a (partially) trained model, and predicts the energy cost of implementing that ML algorithm on the platform? Does this require an actual implementation, or just a well-designed energy prediction function?**

A: The expected output of the process is executable machine learning code. This includes all components necessary for compilation, optimization, and runtime. The program encourages innovation in conversion techniques optimized for each unique platform. Truly hardware-agnostic algorithms are the long-term goal, but within the scope of this program DARPA seeks significant gains from allowing teams to develop algorithms based on their own architecture. The program seeks to “map ML efficiency to physics.” While a well-designed energy prediction function is valuable, implementing the full conversion process is necessary

to empirically validate the energy predictions and demonstrate the feasibility of energy ML in practice. A functional implementation would allow for the collection of data on real-world energy usage.

5. Why and how is MLB-Linpack related to the lower bound of energy? Linpack is mainly used for solving system of linear equations, but ML does not need it.

A: MLB-Linpack is the theoretical lower bound of the energy required to perform an optimization. In the context of ML2P, it is the metric used to evaluate the energy efficiency of machine learning computations, while maintaining a specific accuracy level.

6. What is the meaning of A and b in the context of machine learning?

A: A and b represent data used in a core linear algebra operation representative of the machine learning workload, such as a matrix-vector product or solving a linear system. The MLB-Linpack metric is calculated on this operation to evaluate the energy efficiency of ML solutions while maintaining a specified accuracy level.

7. Regarding energy metering baselines: is wall-plug or node-level measurement acceptable as the standard, with optional finer-grained breakdowns when available? For cases where direct measurement isn't feasible (such as hosted services), would pre-documented energy estimates with clearly stated uncertainty be acceptable?

A: The solicitation calls for detailed hardware and power measurement infrastructure, and that the power measurement hardware should be capable of capturing data with sufficient granularity and accuracy to be considered forensically sound and reliable. While the solicitation does not explicitly specify wall-plug or node-level, the requirement for forensic soundness suggests node-level or finer-grained measurement could be an option. For cases where direct measurement is not feasible, pre-documented energy estimates with clearly stated uncertainty might be acceptable but should be justified in your proposal.

8. What should replicability packages for test and evaluation include to ensure government verification of results?

A: Replicability packages for test and evaluation should include, but are not limited to:

- i. Detailed documentation of the experimental setup.
- ii. Hardware specifications and configurations.
- iii. Software versions and dependencies.
- iv. Data sets used.
- v. Step-by-step instructions for replicating the experiments.
- vi. Power measurement hardware details.
- vii. Uncertainty analysis for all measurements (This is supplementary knowledge).

9. For MLB-Linpack, should we expect to exercise both training and inference on every processing element, or are there reasonable exceptions?

A: The solicitation states the implementation of the MLB-Linpack metric described in Figure 8 for both training and inference (Source: ML2P PS Amendment 01.docx). The goal is to establish a lower bound (LB) metric for assessing the performance of ML2P. The solicitation doesn't explicitly mention exceptions, but reasonable exceptions could exist and should be well-justified within the proposal.

OPEN SOURCE, LICENSING, AND DELIVERABLES

1. Must all software be open sourced? What are the circumstances in which open sourcing the software will not be required?

A: Yes, open-source publication is expected. DARPA strongly discourages the submission of technical solutions that offer restrictive licensing, and per section 3.3.1 of the solicitation “Restrictive licensing within proposed technical solutions will be found to be a significant weakness during the scientific review process.” The key transition objective of the program is to make ML2P software the gold standard for ML construction and simulation of power usage and trade-offs. Performers will be required to publish documentation, algorithms, code, and tutorials they will develop and generate under ML2P awards, as open-source (e.g., the MIT license is strongly preferred) to existing ML repository sites (e.g., scikit-learn) and, when available, the forthcoming DARPA GitHub page, in addition to publishing in conferences (e.g., NeurIPS) and peer-reviewed journals (e.g., IEEE).

2. If we are proposing an approach that runs on an edge device (such as Xilinx ZCU 102 FPGA) with training on a separate workstation/server (such as NVIDIA 8* H200), are we expected to send both the edge device and workstation/server to the T&E team, or just the workstation/server?

A: The T&E team's role is to replicate and independently validate all results from each Development team, including the lower bound for each performer. Therefore, the delivery of all hardware components used for processing and power measurement is essential for the T&E team to perform accurate and comprehensive evaluations.

3. Is there a preferred format or schema for Energy Semantics artifacts so teams can exchange and validate them consistently?

A: The program encourages the use of first-order predicates for a formal representation of energy semantics for ML designed for machine readability. A specific format is not mandated; however, consistency and machine-readability are critical.

- 4. How should Pareto results be reported—per task and hardware, aggregated across multiple tasks/hardware combinations, or both? Is there a recommended format for the data or visualizations?**

A: The solicitation explores a set of multi-objective functions (e.g., a Pareto Frontier), balancing the trade-off of joules with performance, covering a range of common ML tasks, such as clustering and classification. Reporting Pareto results both per task/hardware and aggregated across multiple combinations would provide a comprehensive view. A specific format is not required, but clear and well-documented data and visualizations are expected.

- 5. For the monthly code drops, should performers include build scripts, test harnesses, environment lockfiles, and optionally container images to enable offline government rebuilds?**

A: Yes, performers should include build scripts, test harnesses, environment lockfiles, and optionally container images to enable offline government rebuilds. The government evaluator needs the ability to compile all delivered source code.

- 6. What license posture do you prefer for new open-source code? Should repository links be included in the monthly reports?**

A: The ML2P program prefers open-source licenses such as the Massachusetts Institute of Technology (MIT) License. See Section 3.7 Anticipated Deliverables to be delivered under the ML2P program.

- 7. Is a Data Management Plan required post award? If so, should it explicitly address energy measurement datasets. We're thinking this may cover: schema, calibration logs, and materials needed for reproducibility.**

A: See Section 3.7 Anticipated Deliverables to be delivered under the ML2P program.

ELIGIBILITY, SUBMISSIONS, AND AWARD INFORMATION

- 1. Do you know if this opportunity (or potential Opp) is solely targeting academia?**

A: This solicitation encourages submissions from all responsible sources capable of satisfying the Government's needs, including large and small businesses, nontraditional defense contractors as defined in 10 U.S.C. § 3014, and research institutions as defined in 15 U.S.C. § 632.

- 2. Are we eligible to join as a performer given that we are a UARC.**

A: UARCs are highly discouraged from proposing. UARCs interested in this solicitation, either as a prime or a subcontractor, should contact the Agency POC listed in the Overview section prior to the proposal (or abstract) due date to discuss potential participation as part of the government team or eligibility as a technical performer.

3. The RFI does not mention any possibility of potential resulting RFPs. Is this true?

A: The ML2P Program Solicitation (DARPA-PS-25-32) is a formal request for submissions and has been published to SAM.gov. Please review the solicitation for abstract and proposal submission information.

4. How are abstracts to be submitted?

A: Attachment A1 (ML2P Abstract Template), published with the solicitation, describes the page limit, format, content and submission requirements for abstracts.

5. Is a non-profit organization eligible to apply?

A: ML2P encourages submissions from all responsible sources capable of satisfying the Government's needs, including large and small businesses, nontraditional defense contractors as defined in 10 U.S.C. § 3014, and research institutions as defined in 15 U.S.C. § 632.

6. It is clear from the BAA that FFRDCs are not encouraged to participate. However, it is unclear if FFRDCs are allowed to collaborate or offer resources to performers.

Examples include students performing ML2P-related work at FFRDC facilities through independent collaborations. Could you clarify whether that is allowed? If this involves using enabling capabilities, should the PI list the FFRDC as collaborator along with the capabilities they plan to use?

A: The solicitation prohibits FFRDCs from providing support either as a prime or subcontractor, unless the DARPA Deputy Director grants a written waiver. If the FFRDC is interested in this solicitation, either as a prime or a subcontractor, they should contact the Agency POC listed in the Overview section prior to the proposal (or abstract) due date to discuss potential participation as part of the government team or eligibility as a technical performer.

7. We are interested in submitting to ML2P and are curious if you have any guidance on the approximate budget that we should propose, as the PS only mentioned total budget for the entire program.

A: In accordance with the solicitation, proposers are strongly encouraged to select a cost point that is commensurate with the scale and complexity of the proposed approach.

8. We would like to confirm the correct due date for the ML2P abstracts.

A: Please visit SAM.gov for all ML2P solicitation due dates. As of Amendment 01, the Abstract due date is October 15, 2025, at 5:00 PM Eastern Time (ET).

9. Can a nontraditional performer propose with zero cost share under the anticipated Other Transaction Agreement?

A: In accordance with the solicitation, the OT agreement will not require cost sharing unless the proposer is a traditional defense contractor who is not working with a non-traditional defense contractor participating in the program to a significant extent.

10. Should we assume that NIST SP 800-171 applies when controlled unclassified information is present, and that public releases (papers, code, datasets) require Public Release Center approval before posting?

A: Yes, NIST SP 800-171 applies when controlled unclassified information is present. Public releases (papers, code, datasets) will likely require Public Release Center approval before posting.

11. For Principal Investigator (PI) meetings, is single attendee representation acceptable for small teams? Would virtual participation be available when needed, and should we plan to follow Fly America Act and Federal Travel Regulation per diem for any in-person travel?

A: The solicitation highly encourages budgeting for attendance of relevant personnel at each PI meeting. Single attendee representation may be acceptable for small teams, but ensure key technical staff are present at technical events. Proposers should plan for in person attendance at each PI meeting and should estimate travel based on organizational travel requirements. The Government highly encourages use the Fly America Act and Federal Travel Regulation per diem for proposed travel.

12. Does ML2P provide guidance on typical award sizes or a preferred budget range per performer by phase, or should we propose costs strictly commensurate with scope without nominal caps?

A: In accordance with the solicitation, Proposers are strongly encouraged to select a cost point that is commensurate with the scale and complexity of the proposed approach. Proposers are reminded DARPA anticipates awarding multiple awards.

13. Is equipment and measurement hardware costs, inclusive of providing a unit for test and evaluation, deemed as allowable direct costs? Are there any thresholds or disposition rules we should plan for?

A: Equipment and measurement hardware costs should be realistic and reasonable; thus, costs should be directly aligned with the technical approach.