Academic Grant Program

Call for Proposals: AI Training and Model Development

About This Call for Proposal (CFP)

NVIDIA has been transforming accelerated computing for more than 25 years. We're defining the next era of scientific computing and supporting top researchers doing compelling, computationally intense work to solve some of the world's most challenging problems. Developer programs support researchers in using our SDKs, frameworks, and web services by providing training and creating online communities to help researchers do their life's work using NVIDIA technologies.

NVIDIA solicits proposals for innovative projects related to generative AI, particularly small and large language models, multimodal models, and diffusion models. Projects leveraging AI models must incorporate NVIDIA models, such as NVIDIA Nemotron™, wherever possible. All others must make extensive use of NVIDIA software and AI distributions, such NVIDIA Megatron-Core, Megatron-LM, NeMo™ microservices, BioNeMo™, or FLARE™. Proposals should be attached to one or more of the following themes.

Techniques for Training, Scaling, and Customizing Gen AI

- **Training:** Explore novel approaches and techniques for training AI models, including improvements in scalability, compute efficiency, data efficiency, agentic capability, or safety, security, and privacy in decentralized and federated learning settings. Develop open, rigorous benchmarks to evaluate NVIDIA Nemotron and other foundation models to identify strengths, limitations, and failure modes of models.
- **Distillation:** Explore new ways for using large pretrained models to jump-start the training of derivative/smaller models or to enhance the abilities of existing models on specific domains with the use of in-domain expert teachers.
- **Fine-Tuning/Post-Training:** Propose novel methods or adaptations of existing methods for post-training pretrained models for increased performance on specific tasks, domains, or languages. Design open, high-quality training data and evaluation benchmarks, aiming to achieve superior predictive performance and/or low computational cost. Research areas might include AI for hardware design and verification, enhanced agentic capabilities, and advanced mathematical reasoning.

Exploring New Approaches and Architectures for Gen AI

- **Multimodality:** Delve into the intersection of vision, language, speech, and other modalities, aiming to develop models that can seamlessly understand or generate content across diverse data modalities, possibly under low-latency constraints and in response to incremental inputs in real time. Research areas may also include federated learning with missing multimodalities.
- **Modularity:** Investigate architectures with modular aspects, such as mixture of expert networks, dynamic models, or model routing systems.

Award Details

Selected principal investigators (PIs) may receive:

> Up to 30,000 H100 80 GB hours or equivalent (max. of eight concurrent GPUs)

Not all projects that meet eligibility requirements will be selected for an award. The final award amount will be determined by the NVIDIA awards panel. GPU hours provided to the PI will expire six months after the award; unused GPU hours will be forfeited. Physical hardware will be shipped to the PI.

Applicant Eligibility

Full-time faculty at accredited academic institutions that award research degrees to Ph.D. students are eligible. Postdocs and graduate students must work with a full-time faculty member to submit on their behalf.

Each person can submit one proposal per quarter, a maximum of four proposals annually. For example, one submission is allowed in the first quarter, January–March. Each individual applicant is eligible to receive one award per calendar year.

Proposal Requirements

Proposals **must** follow the proposal template and should not exceed **four pages**, not including appendices.

Expectations of Recipients

Award recipients should make reasonable efforts to acknowledge the support of NVIDIA Corporation and reference how specific hardware and software contributed to project results. Recipients will inform NVIDIA of publications, presentations, open-source code and data releases, and speaking engagements that reference the supported project via the NVIDIA academic grant portal. Failure to report in the portal will influence future award selection.

Please review NVIDIA Academic Grant Program terms and conditions.

NVIDIA