AI Safety Fund Biosecurity and Cybersecurity Q&A

Published 9 December 2024

General Questions

Q1. I am currently in the process of founding my organization, and it is not yet incorporated. Can I still apply for funding? At the moment, I am not able to create an account in the grantee portal without a Tax ID Number.

Applicants that do not have a Tax ID Number can still apply for funding. A unique ID number is required to register in the FLUXX application portal. In place of a Tax ID Number, please enter "P" and your phone number. The Tax ID Number can be updated once it has been received.

As an alternative, if your organization is being established as an independent nonprofit organization, using a fiscal sponsor might be a practical solution.

Q2. Can a proposal be multi-PI?

Yes, proposals can include multiple PIs. Please designate a lead PI on your application, and list collaborators as Partners in your application. You can add as many Partners as needed for your research team.

Q3. Can a proposal cover more than one topic listed in the RFP?

Proposals can address one or more of the subtopics outlined in the RFP; we will also accept proposals that cover other relevant areas of research addressing AI safety risks in Biosecurity or Cybersecurity.

Q4. Does the $600k budget limit include indirect costs, or is it only for direct costs?

All costs (direct and indirect) listed in the budget associated with the proposed research should not exceed $600k. Please review the AISF Indirect Costs Policy and the AISF Capital Expense Policy for more information about budget requirements.

Q5. Are you open to backing prize challenges like the ML Model Attribution Challenge 2 (mlmac.io), aimed at attributing AI agents?

Eligible candidates for funding from the AISF include independent researchers affiliated with academic institutions, research institutions, NGOs, and social enterprises across the globe that aim to promote the safe and responsible development of frontier models by testing, evaluating, and/or addressing safety and security risks. The AISF does not provide funding for management of, or regranting under challenges, fellowship programs, training and development programs, or similar initiatives.

Q6. Will it be OK to have industrial partners get involved in this proposal and will they be given funding directly? In other words, would sub-contracting be allowed?

Applicants can include partnerships in their proposals with for profit or non-profit entities to accomplish research outcomes. The AISF will take into consideration each proposal on a case-by-case basis to evaluate the nature of proposed partnerships and may request additional information during the evaluation period to understand the nature and benefit of proposed partnership arrangements. Proposed subcontractors must be included in the proposal and budget as Partners with explanation of their role and contributions to project outcomes. Any subcontracting, outside of what is proposed in a grant application, will require Meridian's prior written consent to engage any person or entity to perform any part of the Deliverables for the project. Applicants should review the Grantee Contract Template referenced in the RFP for further clarification.

Q7. Regarding the Grantee Budget AISF form: for the Budget Narrative section, should applicants provide a detailed Cost Justification?

Detailed cost justification is not required. Applicants are encouraged to provide sufficient detail in the budget narrative to address assumptions and reduce the need for clarifying questions during the proposal evaluation process.

Q8. Can an organization submit multiple proposals?

Yes, applicants can submit multiple proposals.

Q9. Is AISF specifically looking for proposals such that all of the work products will be distributed to AI Labs, researchers, AISIs, 3rd party organizations, and the public? Or, if it is acceptable for proposals that are intended for only a subset of those entities, which subsets are "fair game"?

The AISF is interested in advancing the science of AI safety broadly, so by default we would assume all research products will be suitable for public dissemination unless there are legitimate information hazards. Furthermore, the AISF will not fund research for consumption by just one lab/entity.

Q10. The submission website asks for contact information (phone, email, etc.) about our institution (university). Shall we leave the contact of the grant office there, or just general department/college contact information? There is also a checkbox for "Individual Applicant"

in the application form. As university professors, shall we apply "on behalf of an organization" or as individuals?

When applicants register in the FLUXX application system, they should register as the PI, and include contact information for the primary point of contact for the proposal. During registration, applicants should list the organization that would be the recipient of funding. The "Individual Applicant" checkbox is for individual researchers not tied to any organization. An "Individual Applicant" would receive and manage the funding.

Q11. Is the budget a factor when considering our proposal? Our proposal will include three faculty members with complementary expertise, and we might request $400k - $500k as the total budget (about $150k each person). Will that be considered a "high range" and decrease the chance of acceptance?

As long as the proposed budget falls within the target range specified in the RFP, it will not be a primary factor in funding decisions. Proposals that meet the eligibility criteria will be evaluated primarily based on the proposal's merit and feasibility.

Q12. Can you confirm if overheads/indirects are an eligible cost on this grant?

The AISF will consider a maximum indirect cost rate of 20% of the total direct costs. This cap ensures that a significant portion of the funding is allocated directly to project activities and outcomes. For more information, please review the AISF Indirect Cost Policy.

Q13. Is there any additional opportunity for submitting additional questions?

The question-and-answer period is closed for the Cybersecurity and Biosecurity RFPs. Additional questions can be submitted to the AISF (AIsafetyfund@meridprime.org), however we cannot guarantee a response to questions received outside of the Q&A period. If a response is provided, the AISF will include an update to responses posted on the website for all applicants to view.

If another RFP is released, applicants are welcome to submit general questions as well as questions related to the scope of a new RFP.

Cybersecurity

Q14. My organization is building AI-enabled cyberdefense tooling to improve security at frontier AI labs. Is this in scope for this RFP?

Yes, this is in scope for the Cybersecurity RFP as it relates to the core objective of "evaluating and improving the safe deployment of AI in cybersecurity contexts." The proposal may be relevant under a few subtopics. For example, it may relate to "uplift studies" if the proposed research investigates how the AI-enabled tools improve defender performance. Or, it may be suitable under "interdisciplinary testing" if the tooling focuses on defense against human-driven threats such as phishing or social engineering.

Q15. Our proposal will study the offensive capability of novel AI systems, following the "realistic evaluations of AI cyberoffense" topic in the RFP. Since the project duration is one year, we might not have sufficient time to study how to defend against our novel offense, and our project may focus only on offense. Will that lead to any ethical concerns during the proposal review since our novel offense may (theoretically) cause damage or harm without known defense?

It is not disqualifying if the research only addresses the offensive capabilities of novel AI systems. We are interested in funding research that advances the field.

Applicants should consider how they will ensure responsible conduct of the proposed research, particularly when it comes to handling potentially hazardous information discovered during the project. If the research findings do carry significant risks, applicants should explain how they will assess withholding of its publication.

Q16. Do the "novel security vulnerabilities" mentioned under "realistic evaluations of AI Cyberoffense" topic include vulnerabilities of AI models in cyberspace? If we submit a proposal on exploiting the vulnerability of AI systems, rather than using AI systems for offense, will that be out-of-scope?

In general, the AISF is happy to accept proposals for research that is outside the specific research objectives outlined in the Cybersecurity RFP. The same is true for the Biosecurity RFP. As long as a proposal supports the overarching goal of the RFP, evaluating and enhancing the safe deployment of AI in cybersecurity contexts, we will consider it. Applicants should review and address the AISF ethics criteria, especially where research involves exploiting vulnerabilities in proprietary frontier models. Specifically, “proposals should demonstrate an ethical approach to all research methodologies, avoiding any practices that may inadvertently mislead or compromise collaborators without their informed consent.” Researchers should not endeavor to “attack” an AI system without the consent of the frontier lab.

Q17. Our project focuses on forecasting implementation timelines for state-actor level (SL5) security measures, driven by both the increasing offensive capabilities of AI systems and their growing value as targets. While your RFP's forecasting category emphasizes threat landscape changes, would you consider forecasting critical security measure implementation timelines as within scope, given these dual pressures?

In general, the AISF is happy to accept proposals for research that is outside the specific research objectives outlined in the Cybersecurity RFP. The same is true for the Biosecurity RFP. As long as a proposal supports the overarching goal of the RFP, evaluating and enhancing the safe deployment of AI in cybersecurity contexts, we will consider it.

Q18. Our project may include components from more than one topic listed on the RFP. For example, we may build the AI offense system with a human interface so our study will also be relevant to the "uplift studies" topic. Do you want us to focus on one topic only during the short one-year period, or is a combination of multiple topics also encouraged?

Applicants should not feel constrained to a single research objective outlined in the RFP. We welcome all proposals that fit the RFP's goal of evaluating and improving the safe deployment of AI in cybersecurity contexts. The same is true for the Biosecurity RFP.

Biosecurity

Q19. My proposal relates to biomedical research but also relates to data and AI safety, which can relate to Cyber Security as well. Would it be OK for me to submit via the route of Biosecurity if my research relates to both Biosecurity and Cybersecurity?

Applicants should submit proposals under one domain. In this case, submitting under the Biosecurity RFP will ensure that the proposal is assigned to a suitable expert for evaluation.

Q20. To what extent would AISF like to see coverage of material that is clearly hazardous? If such material is included, is it incumbent upon the proposers themselves to outline procedures for safe distribution of work output to relevant users?

It is incumbent on applicants to demonstrate that they have a keen awareness of the implications of their research and have taken into consideration safety measures to mitigate harm from any potentially hazardous information or ethical harms that result from their work. The Frontier Model Forum (the FMF), a partner of the AISF, is developing policies and guidelines for safe distribution of information. We expect the policy and guidance to be finalized prior to the completion of any research funded by the AISF for the Biosecurity and Cybersecurity RFPs and will advise grantees to manage research outcomes under the advice of the FMF.