



Request for Proposals: AI Safety Grant Program

Our grant program funds projects that aim to reduce existential risks from AI, focusing on four underexplored areas that could have significant impact on AGI safety and the future of humanity. We provide funding to support high-risk, high-reward initiatives that develop human capabilities, strengthen AI safety frameworks, strengthen both human capabilities and the cooperation architectures between and among humans and AI systems.

Grant Size

Annual funding range: \$4.5-5.5M, split across our different focus areas.

Focus Areas

We are interested in funding projects that address one or more of the following four areas:

Automating Research and Forecasting

- Scaling AI-enabled research to support safe AGI development
- Scaling efficient forecasting methods relevant for safe AGI
- Other approaches in this area

Neurotech to Integrate with or Compete Against AGI

- Brain-Computer Interfaces (BCIs) to enhance human cognition or facilitate human-AGI collaboration
- Whole Brain Emulations (WBEs) as human-like intelligences that are interpretable and alignable
- Lo-fi emulations using behavioral and neural data with deep learning
- Other approaches in this area

Security Technologies for Securing (AI) Systems

- Implementations of computer security techniques (POLA, SeL4-inspired systems, hardened hardware)
- Automated red-teaming, vulnerability discovery
- Cryptographic techniques for trustworthy coordination architectures
- Other concrete approaches in this area

Safe Multi-Agent Scenarios

- Game theory addressing interactions between multiple humans, AIs, or AGIs
- Avoiding collusion and deception, encouraging positive-sum dynamics
- Principal-agent problems, Active Inference agents
- Other concrete approaches in this area

We encourage proposals that integrate approaches which span across multiple focus areas.

Why We Fund This Program

In response to shortening AGI timelines, we focus on funding high-risk, high-reward projects that advance AI safety and mitigate existential risks. While many funders focus on aligning AI systems, our program emphasizes strengthening human capabilities through neurotechnologies, supporting automated research, and securing cooperation architectures through computer security and multi-agent scenario analysis.

Eligibility and Application Process

Eligibility

Open to individuals, teams, and organizations with projects aligned to one or more of the focus areas.

Application Process

- **Fill out the [application form](#):** (includes personal details, project details, and a request for references).
- **Peer Review:** All proposals will first undergo an initial peer review process to assess whether the application is within scope and meets the basic eligibility requirements. This ensures that the project aligns with the four focus areas of our grant program.
- **Technical Advisors & Screening Interview:** If your proposal passes the peer review, it will be sent to at least three technical advisors for further evaluation and insights. We will also invite you for a short, 15-minute screening interview to

discuss your project. Both the review by the advisors and the interview will happen concurrently.

- **Technical Discussions:** Any technical questions or clarifications will be addressed via email during the review process.

Decision Timeline: Decisions are made within 8-12 weeks of the application deadline.

Quarterly Application Deadlines:

- March 31
- June 30
- September 30
- December 31

Feedback

Unfortunately, due to the scale of applications we receive, we are unable to provide personal feedback to the majority of applicants.

Evaluation Criteria

Proposals will be evaluated based on the following criteria:

- **Impact on Reducing Existential AI Risks:** The extent to which the project can reduce existential risks associated with AI, focusing on achieving significant advancements in AI safety within short timelines.
- **Feasibility within Short AGI Timelines:** The project's ability to achieve meaningful progress within the anticipated short timeframes for AGI development. We prioritize projects that can demonstrate concrete milestones and deliverables in the near term (1-3 years).
- **Alignment with Program Focus Areas:** The degree to which the project addresses one or more of the program's core focus areas, including neurotechnology, multi-agent scenarios, security technologies, and methods for forecasting AI development. Proposals should prioritize underexplored approaches with potential for rapid impact.
- **Capability to Execute:** The qualifications, experience, and resources of the project team to successfully carry out the proposed work. Strong teams with proven expertise in AI safety or related fields will be prioritized.
- **High-Risk, High-Reward Potential:** The level of risk involved in the project, balanced with the potential for substantial, transformative impact on the future of AI safety. We encourage speculative, high-risk projects with the potential to drive significant change if successful.

Grant Size

Previous grants have ranged from \$10,000 to over \$300,000. We don't have a set minimum or maximum amount. We encourage applicants to propose the amount they believe is necessary, along with a moderate or ambitious budget that aligns with project goals and scope.

Multi-Year Funding

We welcome both short-term and multi-year projects. For all projects, we expect a comprehensive timeline and funding breakdown in your application. For long-duration and high-budget projects, funding may be provided in tranches, contingent upon the successful completion of prior phases.

Financial Guidelines

- **Payment Structure:** Grants are typically paid in one lump sum. However, for larger projects spanning multiple years, payments may be made in tranches, with each subsequent tranche contingent upon the successful completion and reporting of previous milestones.
 - **Overhead Policy:** Depending on your organization's nature, we may fund overhead costs. Overheads must not exceed 10% of the direct research costs, and must directly support the activity funded by the grant.
 - **Tax Responsibility:** Tax implications vary by jurisdiction. We recommend consulting a tax professional for clarification.
-

Funding Conditions

- **Due diligence:** Applicants must pass our due diligence process, which includes confirming your connections to Foresight Institute, sharing any ongoing criminal proceedings, bankruptcy, etc.; and sharing the following documents:
 - Detailed itemized budget
 - Final agreed project plan
 - Organizational documents (if applicable)
 - Jurisdiction of incorporation
 - List of directors/trustees/board members
 - Recent financial statements (e.g., tax returns, annual reports)
 - Form W-9 (for US entities)

- **Reporting and Accountability:** Grant recipients must submit bi-annual progress reports at regular intervals, detailing how funds are used and the outcomes achieved.
 - These include sharing the milestones agreed upon at the time of the grant and comparing them to your progress.
 - If this grant spans over multiple years and is a large grant, we may choose to fund it in tranches. If so, receiving tranches will be conditional upon using the grant funds wisely and achieving milestones.
-

Media Policy

Media Sharing and Publicity: By accepting funding, grantees agree to the following terms related to media and publicity:

- **Project Listing:** We may list your project on our website and include it in other relevant public communications.
- **Social Media:** Foresight may share details of your project on our official social media channels, including Twitter, LinkedIn, and Facebook.
- **Newsletter:** Your project may be included in Foresight's monthly newsletter.
- **Content Shared:** The typical content shared includes:
 - Grantee's name, affiliation, and headshot
 - Project title and brief project summary
 - Relevant milestones or achievements (link to a preprint / paper, website, etc)

Grantee Consent: We ask for approval from grantees for the use of the above content. If you have any concerns or preferences about media sharing, please communicate them at the start of the project.

Confidentiality and Restrictions: If your project involves confidential aspects that cannot be publicly shared, please inform us. We will respect these confidentiality requirements in all communications.

Opt-Out Option: Grantees may opt out of specific media channels. If you prefer not to have your project shared on our website, social media, or newsletter, please let us know.

What We Do Not Fund

We do not fund projects outside the four specified areas of AI safety outlined above.

Common exclusions include:

- General AI alignment research
 - Application-based AI technology proposals (e.g., AI for healthcare, medical research, education, or business tools)
 - Projects in non-AI safety fields
-

Conclusion

We encourage bold and innovative projects that tackle existential risks from AI. If your project aligns with our funding focus and you're passionate about advancing AI safety, we welcome your application. We are particularly interested in projects that engage in interdisciplinary research and have the potential to drive significant change in underexplored areas of AI safety.

Apply

Please complete the [application form](#) to apply.

If you have questions please contact grants@foresight.org.