



## **AI Safety Grant Program**

This program funds projects that aim to reduce existential risks from advanced AI. We focus on work within four overarching areas:

1. Security technologies for AI-relevant systems
2. Neurotechnology to enhance or compete with AGI
3. Safe multi-agent scenarios involving humans and AIs
4. Automating research and forecasting

**We strongly prioritize proposals that address specific directions within these areas, as outlined in our active Requests for Proposals (RFPs) webpages ([1](#), [2](#), [3](#), [4](#)).**

General proposals are considered on a case-by-case basis, with a far higher standard for selection. We do not recommend you apply with a general proposal.

---

## **Grant Size**

### **Grant Size**

We award \$4.5–5.5M in total funding annually across these areas. Individual grants typically range from \$10,000 to over \$300,000. We do not set a minimum or maximum; applicants should request what they believe is appropriate and justified.

### **Multi-Year Funding**

We welcome both short-term and multi-year projects. For longer or higher-budget work, funding may be provided in tranches based on milestone completion.

## What We Fund

We fund projects that reduce existential risks from advanced AI by targeting specific directions within four core areas. Proposals that engage directly with one of the areas below — as outlined in our active RFPs — are prioritised.

### **Security Technologies for AI-Relevant Systems**

We support projects that secure critical infrastructure and advanced AI systems. This includes automated red-teaming, scalable formal verification, cryptographic coordination tools, and decentralised or privacy-preserving compute infrastructure. We prioritise approaches that can scale with frontier AI capabilities.

### **Neurotechnology to Enhance or Compete with AGI**

We fund work that uses neuroscience and neurotech to make AI systems more interpretable or improve human capabilities. This includes brain-aligned models, lo-fi emulations, secure brain-computer interfaces, and biologically inspired architectures that improve alignment, robustness, or human-AI collaboration.

### **Safe Multi-Agent Scenarios**

We fund work that enables safe, transparent, and cooperative interaction between multiple AIs and humans. This includes systems to detect collusion or deception, tools for coordination and consensus-building, and AI agents designed for pro-social interaction. We're especially interested in real-world testing and simulation of multi-agent safety risks.

### **Automating Research and Forecasting**

We support open-source tools that automate scientific research workflows — such as literature review, hypothesis generation, and experiment design — and AI systems that improve forecasting of technological and global risks. We're particularly interested in research assistants, continuous forecasting platforms, and collaborative modelling systems that enhance human reasoning and decision-making.

---

## 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 in-house review to assess scope and fit. This ensures that the project aligns with the four focus areas of our grant program.
- **Technical Advisors:** If your proposal passes the peer review, it will be sent to technical advisors for further input.
- **Screening Interview and Technical Discussions:** We may invite you for a short, 15-minute screening interview to discuss your project. 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 our four core focus areas: automating research and

forecasting, neurotechnology, security technologies, and safe multi-agent scenarios. We strongly prioritise proposals that directly address specific directions outlined in our RFPs.

- **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.
- 

## 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](mailto:grants@foresight.org).