CAISI
RESEARCH PROGRAM
AT CIFAR

CALL FOR PROPOSALS:

Catalyst Projects on Cybersecurity and Agentic AI in AI Safety

CIFAR

Funded by the Government of Canada | Canada

Table of Contents

Opportunity 3

Areas of Interest 4

- Agentic AI & Autonomous Systems Control 4
- Cybersecurity, Open-Weight Deployment & Systemic Risks 5

Eligibility & Funding 6

- Stream A. Canadian Project Stream 6
- Stream B. Partnered Project Stream 6

Application Process 7

- Timeline 7

Proposal Requirements 8

- Evaluation Criteria 9
- Use of Generative AI 9
- Reporting Requirements 9

Opportunity

The Canadian AI Safety Institute (CAISI) Research Program at CIFAR is pleased to launch a call for proposals for Catalyst Projects that advance fundamental and applied research into AI Safety. Catalyst Awards fund high-risk, high-reward research and exceptionally creative proposals with the potential for broad impact in the field of AI safety.

The CAISI Research Program at CIFAR is the independent scientific research engine of CAISI, leading multidisciplinary research to tackle the safety challenges posed by advanced AI systems. As a core component of the government's broader Canadian AI Safety Institute, the program leverages Canada's robust AI research ecosystem to advance critical knowledge into the science of AI safety.

This call will fund two streams of research: **Stream A**, for Canada-based researchers, and **Stream B**, for eligible international projects. Both streams will fund the salary, benefits and travel expenses of a postdoctoral researcher for up to two years to support projects related to the safety and societal risks of agentic AI and cybersecurity.

For Stream B, we are pleased to partner with international organizations in Germany and France to co-fund innovative projects in AI safety between either Canadian and German researchers or Canadian and French researchers.

CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

The CISPA Helmholtz Center for Information Security is one of Germany's national Big Science institution within the Helmholtz Association, and one of the world's leading research center in the field of cybersecurity and trustworthy AI.

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Inria

Inria is the French National Institute for Research in Digital Science and Technology and supports the French government in national research and innovation strategies in the digital field, acting as Digital Programs Agency. Inria has launched the AI Evaluation program, a research program under the France 2030 initiative, whose purpose is to develop the scientific roadmap for the French Institute for AI evaluation and security (INESIA).

Areas of Interest

This call centers on two critical and closely-linked challenges in AI safety: agentic AI and cybersecurity. We invite proposals spanning technical and sociotechnical approaches, including proposals that address both thematics. We are particularly interested in the following topics:

Agentic AI & Autonomous Systems Control

- **Effective Human Oversight:** Develop frameworks and metrics that treat human oversight as a testable, verifiable requirement across autonomous agentic workflows. Projects might address:
  - Measures of oversight effectiveness
  - Human-factors evidence
  - Strategies for circumventing ‘metrics gaming’
- **Control of Highly Autonomous Systems:** Developing robust methods, containment protocols, and fail-safe mechanisms for enforcing and retaining control over agentic systems, including when an agent is not cooperating. Projects might address:
  - Technical mechanisms (eg. containment, sandboxing, guardian models)
  - Non-technical mechanisms (governance, regulatory & licensing models, incident reporting)
  - Lessons from other safety-critical fields, such as nuclear power or aviation

- **Model Transparency and Interpretability:** Developing methods that make agentic reasoning and decisions auditable and governable. Proposals might address:
  - Decision trails and reasoning traces for AI agents
  - Interpretability and monitoring mechanisms
  - Sociotechnical mechanisms for achieving transparency (auditing) and trade-offs with privacy, security, and IP
  - Safety challenges in high-stake domains, such as health, finance, and public services

Cybersecurity, Open-Weight Deployment & Systemic Risks

- **Cybersecurity & Systemic Threat Mitigation:** Solutions to counter novel cyber vulnerabilities, agentic threats, and systemic risks associated with advanced AI deployments. Proposals might address the following, or other related topics:
  - Detection and monitoring for agent behaviour
  - Security evaluation and red-teaming
  - Exploitation of trust through human-agent relationships
  - Systemic risks (concentration risk, critical infrastructure, information ecosystem integrity)
  - Standards and assurances for mitigating cybersecurity vulnerabilities
- **Open-Weight Model Safety Tools:** Tools and frameworks that lower barriers to the safe, secure deployment of open-weight models, ensuring robust safeguards and risk controls. Topics may include:
  - Measurement and evaluation protocols
  - Safety benchmarks
  - Global standards and policy requirements

Eligibility & Funding

Stream A. Canadian Project Stream

Projects may be submitted by 1 or 2 Principal Investigators (PIs). One of the PIs must hold a faculty affiliation at a Canadian university.

This stream will provide funding to the PI to provide $80,000 CAD for the salary and benefits for a postdoctoral fellow for one year. An additional $20,000 is available for each project to support benefits and the direct costs of research, including computing costs and travel to academic conferences to present the work.

Stream B. Partnered Project Stream

Projects must be submitted by at least 2 Principal Investigators (PIs) to fund either Franco-Canadian or German-Canadian collaborations.

One of the PIs must have a faculty affiliation at a Canadian university.

The second PI must meet one of the following criteria:

- **Inria:** A PI from France who is a member of an Inria research team.
- **CISPA:** A PI from Germany who is a faculty member at CISPA.

This program will provide funding to the PIs for the salary and benefits of a postdoctoral researcher for up to two years. The funding for a postdoctoral researcher based in Germany or France will be set according to the standard rate of the partner organization, up to a maximum of $150,000 CAD per year.

Application Process

Timeline

Please upload Catalyst Project proposals to CIFAR's SurveyMonkey Apply portal.

The application will launch on the portal on **November 23, 2026** and close on **January 4, 2027 (11:59, Anywhere on Earth)**. Please ensure to press SUBMIT.

Please contact ai@cifar.ca if you have any questions or require additional information.

- **Applications open:** November 23, 2026
- **Application closes:** January 4, 2027
- **Decisions expected:** February 2027
- **Funding administered:** by March 31, 2027
- **Projects start:** between March and October 2027

CIFAR's SurveyMonkey Apply portal

Proposal Requirements

Please submit a PDF containing the following information in either English or French:

- Title of proposal
- Name, affiliation(s), and email address of PI(s) submitting the proposal
- **Project Description (7 pages maximum, excluding the list of references):**
  - Project Objectives, Methodology/Theoretical Approach, Expected Outcomes, Timelines, and Milestones
  - A description that addresses how the project is original and significant and its potential societal impact (e.g., informing policy, mitigating harm, enhancing public trust, including equity-deserving groups)
  - If needed, description of the computing resources required for the project and the plan to access them
  - A description of how this project addresses priority research areas of the call and of the CAISI Research Program at CIFAR
- Background on the team: What expertise does this team bring to the project that is essential to move the project forward? (max ½ page)
- Proposed start-date (between March and October 2027)
- Name, affiliation(s), and email address of any confirmed collaborators
- Current CVs for all applicants, collaborators, and known trainees

Evaluation Criteria

Proposals will be evaluated by the CAISI Research Council and external experts according to the following criteria:

- Originality and significance of the proposal: 35%
- Potential societal impact of the proposed research: 25%
- Research excellence of the applicants: 15%
- Feasibility of project (timeline, resources available, expertise of team): 15%
- Relation to priority areas of the call for proposals: 10%

Use of Generative AI

Applicants may use generative AI in developing their applications, but you must fully verify and take responsibility for the accuracy, completeness, and relevance of all content generated by AI. The accountability for the integrity and quality of the final application remains solely with you, the applicant.

For the purpose of all applications, CIFAR's policy is aligned with the guidance provided by the Canadian federal research agencies regarding the use of generative AI.

Reporting Requirements

- Successful recipients will be required to provide a report to CIFAR after the first year and within 45 days of the conclusion of the project that highlights the major findings and impacts of the research.
- Successful recipients will be expected to attend and may be invited to present at the annual CAISI Research Program Meeting (travel expenses will be covered), held in Fall 2027.