CAISI Research Program at CIFAR

Canada CIFAR

CALL FOR PROPOSALS

Catalyst Projects on Sociotechnical Considerations in AI Safety

Overview

The Canadian AI Safety Institute (CAISI) Research Program at CIFAR is pleased to launch a call for Catalyst Project proposals for projects on sociotechnical considerations in artificial intelligence (AI) safety. We invite applications led by researchers in the social sciences and humanities (SSH) whose work addresses the pressing social, ethical and governance considerations of AI safety.

The CAISI Research Program at CIFAR is independently leading Canadian, 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.

Catalyst Projects are novel and exceptionally creative research projects with the potential for broad impact in the field of AI safety. In this call for Projects, funding will be up to $70,000 per year for up to two years. By advancing research in this important and urgent area, we aim to build on Canada's existing strengths in sociotechnical AI safety research and to bring together researchers from various disciplines and build a community of SSH researchers working on AI safety across Canada.

We understand AI safety to be a broad-ranging field that includes both technical and sociotechnical dimensions. Our emphasis is on the risk mitigation of advanced AI systems — leading-edge frontier models — and the social, political, ethical, legal and economic implications of such systems outlined in the final International AI Safety Report 2025 (released January 2025). We are particularly interested in projects that address the priority research areas for CAISI, which include synthetic content; AI alignment; advancing knowledge on the design, development, and deployment of safer AI systems; studying the properties of complex AI systems and their real-world impacts; and improving risk governance (assessment, assurance, oversight).

As AI systems grow in complexity, capability, and influence, ensuring their alignment with human values and norms, social institutions, and long-term societal goals becomes critical. While technical research plays a central role in developing safe and robust AI systems, SSH disciplines — particularly the qualitative, interpretive, and humanistic disciplines — offer essential insights into human behaviour, social systems, moral reasoning, institutional design, historical precedents, and public accountability — all of which are vital for designing and governing safe AI.

The inputs and outputs of advanced AI systems are increasingly moving into realms of complex cultural data, requiring contextual, interpretive judgment — the precise kinds of interpretive methodologies that humanists and social scientists are best poised to apply. Mitigating the risks of these systems also requires innovative approaches that acknowledge and address the fact that the technology is already embedded within and interacting with society.

SSH scholars have an important role to play in building better, safer AI systems — including AI architectures, benchmarks, evaluation frameworks, and training strategies — that take this cultural and social complexity into account and that have the potential to advance AI's ability to solve challenges, enhance human potential, and bring about positive changes to society and humanity. SSH scholars also have a role to play in critically engaging the concepts, histories, and practices of safety.

Areas of Interest

Governance, institutions & public engagement

- What norms, institutions, or multilateral arrangements are needed to oversee and coordinate the safe development of frontier AI? How do power and social dynamics (e.g., among industry, government, and public; between superpowers and other countries) influence them?
- How to define safety and risk from societal perspectives? How do public perceptions of AI safety—as either an enabler of trust or a barrier to adoption — ultimately influence the speed and scale of AI integration into society?
- How can norms, institutions, or democratic governance mechanisms guide safe development, deployment, adoption, and oversight of advanced AI systems (e.g., system-level security)?
- What mechanisms ensure public engagement and democratic input in AI safety governance, including but not limited to enhancing AI literacy across Canada?
- What AI governance and policy structures enable AI safety and ensure the benefits of AI are distributed equally?
- How can global inequalities be addressed in the governance of advanced AI systems?

Risk, uncertainty & societal foresight

- What societal impacts, including but not limited to those on individual well-being, social interactions/relationships/structures, meaning and availability of work, and labour market and economy, might arise from new and emerging capabilities of frontier AI, and how can we preempt or mitigate associated risks?
- What kinds of innovative methodologies can be used for conducting research into AI safety and developing safer AI systems, benchmarks, or evaluation approaches (e.g., foresight, qualitative methodologies, other social sciences/humanities methods)?
- How might lessons from historical comparisons (e.g., nuclear governance, aviation, biotechnology, financial regulation) inform the management and deployment of transformative technologies like AI?
- What can we learn from the history of our rhetoric, ideas, and practices around safety, risk, and AI?

Ethics & value alignment

- What ethical and legal frameworks should underpin the design, deployment, evaluation, and monitoring of AI (e.g., human-centered evaluations), particularly in high-stakes or uncertain contexts?
- How can SSH research contribute to understanding and addressing potential misalignments between AI objectives and human values (e.g., using approaches from philosophy or psychology)?
- What does value alignment mean in diverse sociocultural contexts, and how can it be operationalized in AI development?
- How do public narratives, media framings, and cultural beliefs shape the perceived legitimacy and trustworthiness of frontier AI?

Generative AI/Synthetic content

- How to collect training data that improves model safety? How to responsibly integrate outputs of models into the web ecosystem? How to audit the impact of AI predictions on algorithmic decisions?
- How is synthetic content spreading in the information environment and what systemic impacts is it having (e.g., on individual and community safety and public trust)? How to mitigate potential misuse (e.g., to commit fraud or sway political opinions)?
- How are digital content transparency (DCT) techniques being adopted and used in the real world and what are their potential systemic impacts?
- How to ensure AI-generated content and human-generated content have provenance information that allows us to distinguish them? How to deploy and govern such provenance mechanisms?
- What are the psychological and societal impacts and implications of human engagement with advanced AI systems, including LLMs? How to mitigate potential harm (e.g., through design features, guardrails, and governance frameworks)?
- How can SSH methodologies and approaches support better, safer human-AI ensembles or interactions?

Creative, innovative topics not explicitly addressed here are welcome.

Eligibility Criteria

- Projects can be submitted by 1 or 2 Principal Investigators (PIs)
- PIs must have a faculty affiliation at a Canadian university
- The primary PI must have expertise in SSH, including but not limited to: sociology, political science, economics, history, philosophy/ethics, legal studies, science & technology studies, or media studies

We strongly encourage collaborations between SSH researchers and technical researchers (e.g., computer scientists). CIFAR can help to support matchmaking across disciplines. Applicants are welcome to reach out to ai@cifar.ca. We will try our best to connect you with researchers from other disciplines.

Application Process & Timeline

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

The application will launch on the portal on **October 3, 2025** and close on **November 14, 2025 (11:59, Anywhere on Earth)**.

Please ensure to press SUBMIT.

- **Applications open:** October 3, 2025
- **Application deadline:** November 14, 2025
- **Decisions expected:** January 2026
- **Projects start:** before March 31, 2026

We anticipate funding up to 6 projects in this cycle.

CIFAR's SurveyMonkey Apply portal

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

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 for CAISI and advances sociotechnical aspects of AI safety
- 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
- **Funding Request and Justification:** Amount requested (up to $70,000 per year for up to two years) and detailed budget breakdown (see eligible expenses below)
- Name, affiliation(s), and email address of any confirmed collaborators
- Current CVs for all applicants, collaborators, and known trainees

Eligible Expenditures

This program will provide funding to the PI(s) up to $70,000 CAD per year for up to two years for the direct costs of the research project, including:

- Salaries, stipends, and benefits of researchers, technicians, and support staff for employees portion of time spent on the project;
- Operations of core research facilities;
- Equipment; scientific collections; computer hardware or software; information databases;
- Direct costs of knowledge creation;
- Professional and technical services;
- Material and supplies;
- Expenditures related to knowledge mobilization (e.g., policy development; tool development; Intellectual Property creation and protection; market study; prototype development);
- Expenditures related to training and capacity building (e.g. costs of students, HQP, trainees, and recruitment);
- Travel costs in line with the Modern Travel Practices of the National Joint Council Travel Directive (NJC);
- Expenditures related to accessing computing power and compute infrastructure, up to 20% of individual project costs.

If applicable, CIFAR will provide support to successful applicants in recruiting highly qualified postdoctoral fellows. Fellows recruited through this program will be invited to be part of CIFAR's AI Safety Postdoctoral Fellows community.

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 — 25%
- Potential societal impact of the proposed research — 35%
- Research excellence of the applicants — 15%
- Feasibility of project (timeline, resources available, expertise of team) — 15%
- Multidisciplinarity of the project team (e.g., SSH and technical) — 10%

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 give a brief virtual presentation of their research findings to the CAISI Research Council during one of its regular meetings.
- Successful recipients will be expected to attend and present at the annual CAISI Research Program Meeting (travel expenses will be covered), held in Fall 2026.

CIFAR's SurveyMonkey Apply portal

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