Suggestions for Generating a Project Proposal for SERI's Summer Research Program

- 1. Learn about global catastrophic and existential risks.
- 2. Learn about what kinds of research projects are good fits for SERI.
- 3. Review lists of potential project ideas, and/or generate your own.
- 4. Implement additional suggestions
- 5. Narrow/refine your ideas through office hours.

1. Learn about global catastrophic and existential risks.

Read <u>this article</u> (also available in audio version--see the button at the top of the article) to learn about major global catastrophic and existential risks. This article provides very useful context for knowing what kinds of projects are good fits for SERI, and for better understanding potential research topics.

Learn about what kinds of research projects are good fits for SERI.

Carefully read and implement the following suggestions:

We are looking for projects that, if successful, would (perhaps very slightly) reduce the likelihood that human civilization will permanently collapse, that humanity will go extinct, or that the potential of humanity will otherwise be lastingly diminished.

In other words, we are looking for projects that are not only relevant to issues such as climate change, nuclear war, AI, and pandemics, but would also help prevent global, existential-level catastrophes from these risks.

We are unfortunately unable to support many applications that do not make a compelling case for their relevance to these specific risks.

This means that, to generate a project idea that is a good fit for SERI, it is typically a good idea to focus on worst-case scenarios: look for research questions that will help humanity avoid worst-case scenarios, either by better understanding these potential catastrophes, or by better understanding potential solutions.

Also, please recall that your research proposal should be under 500 words and include the following:

• A research question

- A methodology--how you will answer your research question
- An argument for why your research project will help reduce risks of global/existential catastrophe
- Your main doubts/uncertainties about your research proposal

3. Review lists of potential project ideas, and/or generate your own.

- See past SERI projects on SERI's website
- Learn more about specific GCRs/existential risks
 - See this reading group syllabus on existential risks & social sciences
 - For more in-depth discussion, see Ord's <u>The Precipice: Existential Risk and the</u> <u>Future of Humanity</u>
- See lists that include potential project ideas:
 - o "Policy and Research Ideas to Reduce Existential Risk"
 - "List of Research Questions that Could Have a Big Social Impact, Ordered by Discipline"
 - The more relevant sections are those on climate studies, China studies, economics, synthetic biology, ML/AI, political science/IR, psychology, public policy, and sociology
 - The sources cited at the bottom may also be useful
 - "Project Ideas in Biosecurity for EAs"
 - (Written in the context of effective altruism but relevant for anyone interested in biosecurity projects)
 - List of SERI project ideas from 2020

4. Implement additional suggestions

(This section is written in particular for applicants who have less experience with research.)

Keep in mind:

- Project proposals are only good fits for SERI if they have a credible path toward contributing something (potentially just a little) to existential or global catastrophic risk reduction--the point is to help solve these pressing problems.
- Chances are, research has already been done on your topic of interest--if you
 familiarize yourself with this research, this will greatly help you find ways you can
 contribute and effectively communicate your research.

More concretely, to avoid common pitfalls, implement the following suggestions:

- As advised in sections 1 and 2, familiarize yourself with what kinds of projects we are looking for, and choose a project topic that is within this scope.
 - We are especially interested in projects that have plausible paths to reducing large-scale, long-term risks, such as by protecting humanity from worst-case catastrophes.
- Choose a project that is feasible for you to execute well within 10 weeks.
 - High-quality, focused research is often more valuable than broad, shallow research.
 - We are rarely excited about projects that propose to carry out shallow overviews of problems when there are already experts on these problems.
 - No, a 10-week project is not enough to (for example) figure out how humanity should govern all of AI.
- Spend some time learning about what research has already been done on your topic of interest.
 - Your research will be better able to make valuable contributions if it builds on previous work (or deliberately takes a different approach) than if it is uninformed by previous research.
 - Check that your proposal has not been done before, or if it has been done before, have a compelling case for how your project will add something new.
 - The materials you read should influence your proposal in at least two ways:
 - You should cite a few of them (e.g. when discussing the background to and rationale for your research proposal) to show us you are familiar with them
 - More importantly, previous research should inform your choice of research question and methodology
 - Recommended approaches to this:
 - Reading the <u>Wikipedia</u> article on your topic of interest, and skimming the references
 - Looking over lists of publications by relevant research organizations, e.g.:
 - On various relevant topics:
 - o FHI's publications
 - o Informal publications on the EA Forum
 - o CSER's publications
 - On nuclear and biological risks:
 - o NTI's resources
 - On Al:
 - Governance:
 - Publications by GovAl and CSET
 - Design:
 - Publications by <u>CHAI</u>, <u>OpenAI</u>, and <u>DeepMind</u> on technical AI topics, as well as informal publications on the <u>AI Alignment Forum</u>
 - On relevant philosophy and economics questions:

GPI's publications

- Many institutions not listed above have relevant publications. To find these, look up your topic on <u>Google Scholar</u> or look up a relevant paper on <u>Connected Papers</u>, then read the abstracts of papers that have relevant-looking titles, and read/skim a few papers that seem most relevant/useful
- Develop a specific methodology for answering your research question.
 - We are looking for methodologies that, if executed well, would provide useful insights into your research question.
 - This usually requires specificity--very vague methodologies often make it difficult for us to be confident that you will carry out research in ways that create valuable insights about your research question.
- Think through, be sure to have, and be sure to clearly communicate a credible path that your project has toward mitigating existential and global catastrophic risks.
 - Just because a project is relevant to existential risks does not mean it has a good chance of reducing these risks. We are looking for projects that are not just relevant, but helpful.
 - o Think through:
 - How might people use your research findings to make decisions that are more successful at reducing existential risks?
 - Who will be making these decisions, and what steps can you take so that your research reaches them?

Examples of these things done well and not so well:

Done well	Done not as well
Within SERI's scope: "Research question: What do climate models imply about the worst-case scenarios that could result from solar geoengineering?"	Not within SERI's scope: "Research question: What impacts is climate change having on health in California?"
Feasible in 10 weeks: "I will create toy models of crisis bargaining in the above context to investigate potential mechanisms for mitigating rapid crisis escalation."	Not feasible in 10 weeks: "Research question: What government policies would best reduce risks of pandemics?"
Informed by initial research about topic: "Since tight feedback loops are missing from existing forecasting tools, I will create and test exercises that will enable people to get rapid feedback on their forecasting skills."	Not informed by initial research about topic: "Research question: What government institutions would better represent future generations?" (Multiple papers that answer this question already exist.)

Specific methodology:

"First, I will conduct quantitative analysis using the Grantham Research Institute's Climate Change Laws of the World. I also intend to log which domestic climate laws refer to a target from 1990-2020. With this preliminary data, I would write a programme to extract useful information, which I will analyze for empirical guidance as to what components are key to focus on in designing a robust mechanism for signalling consensus around long-term challenges."

No specific methodology:

"I will use my studies of economics to answer my research question."

Clear path to impact:

"I believe this project has the potential to help close a significant vulnerability in global biosecurity defences at a critical juncture, by proposing DNA screening design specifications that address the problem discussed earlier. This would reduce the risk of release of a synthetic pandemic-causing agent--an existential threat."

No clear path to impact:

"How can the cost of accelerated carbon emissions reductions be justified by the social cost of carbon?"

5. Narrow/refine your ideas through office hours.

Here is our office hours schedule.