

Summary	2
BERI's work	2
Past successes	3
Proposed activities	4
BERI core operations	4
CHAI collaboration	4
Additional work with additional funding	5
Trialing Policy	5
Budget	6
Expected outcomes	6

Summary

The mission of the Berkeley Existential Risk Initiative (BERI) is to improve human civilization's long-term prospects for survival and flourishing. We provide administrative, logistical, technical, and occasionally financial support to researchers and other advocates working to reduce existential risks (x-risks).¹ We are seeking \$403,890 to support our core team and hire staff for a collaboration with the Center for Human-Compatible AI (CHAI), and possibly an additional \$245,810 to support scaling up our activities. We expect to support several major AI Safety research organizations over the next year, in particular CHAI and the Future of Humanity Institute (FHI), as well as individual researchers.

BERI's work

At the most basic level, BERI wants to reduce barriers to additional x-risk research and advocacy. Because many researchers work at universities, we expect to primarily collaborate with universities, providing rapid, flexible assistance wherever BERI can help.

BERI currently acts as a vehicle for hiring contractors and part-time employees for sporadic tasks in support of x-risk research and activities. Potential partners approach us with known gaps we can fill for them, and we discuss the feasibility of BERI doing so. All tasks we take on are first vetted by BERI's Executive Director for their impact and replaceability, in consultation with our advisors. As time progresses, BERI expects to accrue a network of contractors and other supporters who can be leveraged in favor of any project we consider worthwhile to engage with. We also expect to gain additional advisors who can inform us of more diverse x-risk interventions suitable to BERI. In our [Proposed Activities section](#) below, we go into detail about the specific types of services we hope to provide in the near future.

There are several advantages to having BERI perform some types of tasks for researchers. As a small non-profit, BERI can easily "try before we buy": with very little bureaucratic overhead, we can trial several contractors on short projects before hiring someone for a longer engagement. Due to our small size, we can also make funding and capacity decisions rapidly, which is especially helpful in time-sensitive situations. This is often significantly more difficult for larger institutions to do. Finally, because BERI has few restrictions on what it can fund, we are highly flexible with regard to the types of assistance we can provide.

In addition to boosting productivity, we hope that the inexpensive assistance BERI provides will help make the difference, to some researchers, between feeling like a typical academic versus feeling a deep sense of support from one's surrounding social context. Support from BERI means someone is going above and beyond to help you, because they believe in what you are

¹ Note that BERI is currently only focusing on supporting people working on AI Safety, because its leaders have experience in that area. However, it may expand to support other x-risk areas in the future.

trying to do. We hope that our actions will not only accelerate the work and professional development of x-risk researchers, but also help to motivate them.

Past successes

BERI was incorporated in February 2017. Since then, most of BERI's efforts have gone towards creating a new organization and discussing potential projects on which our help has been requested. However, there are already a few projects that BERI contractors have assisted AI safety researchers with:

- **Event coordination with CHAI.** This year, CHAI hosted its first conference, but because it was arranging the conference on short notice and after business hours, U.C. Berkeley's staff were unavailable to assist with setup. BERI quickly stepped in, hiring an event coordinator to assist CHAI. A Berkeley staff member's testimonial is in this footnote.²
- **Web development with CHAI.** BERI has been lending a contracted web developer's services to assist in maintaining CHAI's website while CHAI works to hire its own contractor through UC Berkeley. BERI's contractor turns around change requests quickly, often on the same day as receiving them, without using up the time of a student researcher; a student's testimonial is in this footnote.³

Furthermore, BERI is already in discussions with several other researchers, advocates, and institutions about the support it could provide. For example, the Future of Humanity Institute at Oxford has expressed interest in collaborating with us, and a few small projects are currently in development with them.⁴

²**Gianna Kone, UC Berkeley; Administrator and Programmatic Assistant to Stuart Russell:** "CHAI's workshop this year fell outside of our regular working hours, and university personnel were not available to help on short notice. Working with BERI's event coordinator Roxanne Heston was extremely helpful to us, as she was able to be there when nobody else could be... If we had an ongoing collaborator like BERI that is familiar with Berkeley's contacts and policies, they could be counted on to fill in gaps like these between department administrative staff availability and the needs of CHAI that happen outside our normal business hours. If BERI continues to foster a relationship with Berkeley faculty and staff, they can help extend the best of Berkeley culture to our visitors and collaborators."

³**Smitha Milli, incoming UC Berkeley CS PhD student:** "I created and used to manage CHAI's website. Since we began collaborating with BERI, now BERI's professional web developer is taking care of it instead. I (or others in CHAI) can just email him a request and he takes care of the rest. He finishes our tasks quickly, usually on the same day, which is much faster than I was able to respond to website requests, and requires almost no oversight. This means less of my attention is spent on maintaining our website, and I can focus even more on developing my expertise in AI and AI safety."

⁴**Andrew Snyder-Beattie, University of Oxford; Director of Research, Future of Humanity Institute:** "I'm excited to see BERI working closely with CHAI, and hope that as its capacity grows, we at Oxford might benefit from working with BERI as well."

Proposed activities

BERI has been operating with one volunteer Executive Director, one part time Operations Manager, and one part time IT Consultant, as well as a few contractors to help with event planning and internal operations. We would like to retain and slightly expand our core staff to better support a variety of future projects that we expect to assist with.

In addition, **we are seeking to hire engineers and other staff to collaborate with the Center for Human-Compatible AI on the development of AI virtual assistant software** and related research efforts. CHAI expects its work to advance our understanding of how to control and align the behavior of advanced AI systems with human values.

Below we describe the roles we would like to add to our team. The cost of these staff can be found in the [Budget section](#) below.

BERI core operations

Our core staff coordinate BERI's new projects and keep BERI running smoothly. Ideally, we would like to be able to commit 1-2 years of pay to our staff.

Below are the roles that we would like to fund:

- **Deputy Director.** A part-time position. This staff member would support our Executive Director, who would like to spend at most one day per week managing BERI, given that he is also a full-time AI Safety researcher. The Deputy Director may replace the Executive Director in the future.
- **Operations Manager.** A part-time position. This staff member would replace our current Operations Manager in overseeing onboarding, payroll, legal compliance, donations, and other general operations tasks.
- **Writer/Editor.** A contract position. This person would help BERI with its written tasks (website upkeep, communication with external parties, internal records and policies, grant proposals, etc.).

CHAI collaboration

CHAI has expressed interest in collaborating with BERI on CHAI's existing research agenda. As a small non-profit, BERI can quickly hire developers and other contractors on task-oriented projects. Often, these are tasks that would not be useful to the career of a young academic researcher (whose incentives are mainly to produce written research), but may be laudable accomplishments for the contractors BERI hires if they are interested in working in industry, where concrete products and outcomes are valued.

Below are the roles we would like to hire or retain to support CHAI, roughly in order of priority:

- **Machine Learning Engineer.** A contract position. This person would assist with open-sourcing CHAI software packages, reproducing existing AI research results in a form that is usable to CHAI, developing real implementations of theoretical algorithms, measuring the relative efficacy of algorithms, and more.
- **Special Projects.** A contract position. This person would be a generalist who could assist with a variety of sporadic tasks, such as researching design firms or identifying ergonomic work set-ups.
- **Event Coordinator (Bay Area only).** A contract position. This person would help researchers organize events, by assisting with finding venues, registration, on-site logistics, and so forth.
- **Latex Typesetter.** A contract position. We would offer researchers this person's services to save them time on formatting their papers and reconciling packages.
- **Python Developer.** A contract position. This person would assist researchers in resolving conflicts between packages and reviewing codebases.
- **Illustrator.** A contract position. This person would provide clean, attractive graphics for researchers' papers or presentations.

More details about these roles can be found on the [jobs page](#) of our website.

Additional work with additional funding

While hiring people for our collaboration with CHAI is our top priority, if we were to receive additional funding, we believe we could relatively easily find opportunities to offer many of those same contractors additional work. Some of that work may be additional work with CHAI, while some of it could be with other researchers in our network.

Our hope is that, eventually, we can have sufficient contract staff available to easily provide researchers with whatever type of research assistance they might need. This could include setting up and funding snacks in the office, helping to book travel, setting up ergonomic work environments, providing grants to sponsor visiting researchers, providing funding for workshops, courses, or conferences that would help develop researchers' mathematical and computer science skills, and essentially any amenity that a company could provide to its staff to boost or reward their productivity.

Trialing Policy

BERI has a standard policy of trialing potential employees and contractors on smaller, lower investment projects, typically for 2 months, before offering them larger projects that require significant onboarding and/or input from BERI or its collaborators. We believe that it's important to hire excellent staff and contractors, and we may need to try a few people per role before finding someone who is capable, reliable, and communicative in the ways that we need. In our budget, we have added several trial periods per role to reflect this. As mentioned above, our

ability to easily trial contractors is one of BERI's strengths relative to university-based research groups, and part of why we think we can do great work.

Budget

We have outlined several possible funding scenarios. Details are provided [here](#). In sum, we estimate that:

- The cost of supporting BERI's core operations for the next year and a half is \$162,490.
- The additional cost of hiring contractors for our collaboration with CHAI for the next year and a half is \$241,400.
- The additional cost of hiring contractors for additional work (our "stretch" goal) for a year and a half is an additional \$245,810.

We have also provided scenarios for shorter and longer periods of time.

Roughly, BERI has approximately \$50,000 available: about \$30,000 in unrestricted funding currently on hand, nearly \$15,000 pledged from donors, and a \$5,000 loan to be paid back in 2017.

Note that, while we feel fairly confident in the total figures of our budget, within the budget for any particular collaboration (e.g., a grant for collaboration with CHAI), we intend to be quite flexible in redistributing funds towards whichever roles seem most valuable for that collaboration.

Although BERI is still quite young as an organization and does not yet have a robust track record⁵, we are seeking funding now because it is difficult to operate an organization that relies primarily on human capacity without being able to guarantee potential contractors that we have the funding to support at least a year of work. It is also difficult to feel justified in committing BERI core staff time to our projects if we ultimately will not have enough funding to complete those projects. Having at least a year and a half of funding secure would help us move forward confidently and quickly with the CHAI collaboration and other projects.

Expected outcomes

BERI has only operated a very short while, which makes it difficult to predict the outcomes we expect over the next year. Additionally, BERI's purpose is in part to reactively handle sporadically arising needs, which, again, makes predicting outcomes challenging. We do expect, however, that our collaborators will end up feeling satisfied and supported, and to the extent that we trust our choice of collaborators, we consider this a useful proxy for success.

BERI intends to keep track of the following information privately (i.e., we will share this information confidentially with funders, but will not post it to our website):

⁵ However, our staff have substantial track records with other organizations. See our [Team](#) page here.

- Requests received for BERI's assistance;
- Our reasons for why we do or do not oblige requests from researchers for our help or collaboration;
- Researcher satisfaction, and metrics to be determined via conversation with researchers (e.g., estimates of hours of their time BERI is able to save, number of interruptions we prevented, or similar); and
- Hours our contractors work

We are also open to tracking other metrics that funders are interested in, if the cost of doing so is not too high (especially to the researchers we collaborate with, whose time and attention we are committed to respecting).