SERI 2020
Summer Undergraduate Research Fellowship
FEATURED PROJECTS
1. A Guide to Engineering Buildings for the Next Pandemic

Arvie Violette | Full-Time Fellow
Mentor: Dr. Milana Trounce, Stanford University
→ Learn More
More than ever, the global population is at risk for a potentially catastrophic infectious disease outbreak. Amid the concerns of engineered microbes, natural pathogens, and bioterrorism, there is a growing demand for nonspecific, preventative interventions, especially those that don’t rely on the agency of individuals.

Imagine if we could integrate disease-preventing technology directly into the buildings we visit everyday. This is the idea behind engineering controls, a type of intervention that uses infrastructure to reduce transmission rates such as antimicrobial surfaces or ventilation systems.
The Guide to Engineering Buildings for the Next Pandemic is an interactive and accessible tool that summarizes the research, implementation, and sustainability behind engineering controls suitable for non-medical buildings. The platform allows you to select qualities about your building and budget in order to recommend specific engineering controls that are most suitable for your project. Behind each engineering control is a comprehensive summary including cost-benefit concerns, methods of implementations, sustainability, and more.
2. Cruxes in Forecasting AI Takeoff Speeds

Jack Ryan | Full-Time Fellow
Mentor: Buck Shlegeris, Machine Intelligence Research Institute
→ Learn More
I.J. Good’s “intelligence explosion” thesis claims that an artificial intelligence that is sufficiently advanced will be able to build a smarter version of itself, which could in turn build an even smarter version, and that this process would eventually lead to levels of intelligence much beyond that of humans. It is dubbed an “explosion” because this recursive self improvement process could happen very quickly in theory.
It seems there is significant reason to believe that artificial intelligence will eventually exceed human capabilities--perhaps via an intelligence explosion--although there is much uncertainty about when, how, and how quickly this will happen.

My project sought to help answer the question: What will the graph of AI capabilities over time look like, and will there really be an “intelligence explosion” leading in “AI foom”? This question is a much debated topic in the AI safety community—some think that the nature of artificial intelligence is bound to give rise to a foom, while others look at the smooth and gradual progress throughout history to conclude AI is most likely to develop smoothly and gradually. The question is also imperative: if proponents of fast-takeoff are right, then AI will be much more dangerous because alignment problems might be less evident to AI developers and AI safety researchers won’t have as much time to test alignment solutions on weaker but analogous AI systems.
To better understand this question, I lay out the relevant facts for predicting AI growth rates, as well as what seem to be the cruxes leading to disagreement among AI forecasters. Lastly, I suggest that we need a more formal model of the intelligence explosion as well as further research into the possibility of an AI overhang so that our predictions can be more informed.
3. A Comparative Analysis of Small States as Models for Existential Risk Mitigation Efforts

Sam Good | Full-Time Fellow
Mentor: Dr. R. James Breiding, S8Nations
→ Learn More
Societies of the modern era are being forced to confront a tsunami of pressing catastrophic risks never before seen in human history. Other research establishes the fact that many solutions and effective preparedness efforts in regards to such risks are increasingly coming from small, adaptive nations. This project highlights the ideas and policies of small nations that are aiding in the mitigation of human existential risks. Specifically, through extensive literature review, interviews, and the production of a podcast series, a number of "Big Lessons from Small Nations" were highlighted, with emphasis on future-minded policies and climate change mitigation efforts. The results show a significantly higher density of such efforts in terms of small nations' policies, innovations, and cultural values. The work and research contributed directly to the success of S8Nations, a Zurich-based think tank focused on the global impacts of small nations.
A COMPARATIVE ANALYSIS OF SMALL...

18% of the world's nations have a futures representative in place

50% of the S8 member nations have a futures representative in place
4. Widening the Framework for Regulation of Dual-Use Research in the Wake of the COVID-19 Pandemic

Harshu Musunuri | Full-Time Fellow
Mentor: Cassidy Nelson, Oxford University
→ Learn More
In the wake of the COVID-19 pandemic, the global scientific community has directed increased resources and attention towards the study of viruses and emergency medical countermeasure development. Our analysis of publicly available data on publications, preprints, and clinical trials coupled with the Global Health Security Index measure of dual-use oversight suggests that some research activities raise dual-use concerns that have gone unaddressed. While most of the research in this domain is essential to advance vaccines, therapeutics, and diagnostics against SARS-CoV-2, improving responses to natural pandemics must not come at the expense of increasing risks of intentional biological threats.
We argue that certain categories of research that receive more attention following a pandemic pose unique dual-use risks that must be addressed by a comprehensive approach involving:

1) a broader definition of dual-use research of concern that captures experimental techniques that could feasibly be translated to harmful pathogens.
2) creation of regulatory frameworks to oversee the funding and publication of dual-use experiments, especially in countries with large biomedical research output and countries where the GHSI has highlighted shortcomings in national dual-use policies.

These policy changes must also be accompanied by stronger social norms among grantmakers and scientists alike, in order to effectively address biosecurity concerns in advance of the next outbreak.
5. Nukes and ‘Red Phones’: A Threat Assessment of Bilateral Hotline Insecurities Between International Powers

Katharine Leede | Full-Time Fellow
Mentor: Dr. Herb Lin, Stanford University
→ Learn More
Nuclear weapons pose an existential threat to the world. Developing technologies such as digital media, artificial intelligence, super-electromagnetic pulses, and anti-satellites weapons can dramatically accelerate escalation in conflicts prone to nuclear exchange. Indeed, the most probable cause of full-scale nuclear war may be the result of miscommunication or miscalculation. As a result, reliable and secure communication networks are indispensable during escalatory crises.

One of the oldest, and most famous, communication lines is the Washington – Moscow hotline, otherwise known as the “red phone.” Since its activation in 1963, an additional 44 bilateral hotline links have been established between countries, 10 of which are top-level, leader-to-leader lines. However, the mere act of establishing a hotline does not guarantee that, during a crisis, reliable and effective communications will occur.
A THREAT ASSESSMENT OF BILATERAL...

FIGURE 2*:

*The hotline between ministers in Figure 2 refers to lower-level political or military ministers. Several countries have also recently established “space hotlines” to improve negotiations in the space conflicts
The purpose of this study is to provide an overview of the ways in which hotline communication can fail in order to answer the question on how they can be improved. Buttressed by case studies, this project will explore a variety of possible impediments to successful hotline communications, including technical, political, psychological, and organizational dimensions. While information on the Washington – Moscow hotline is present in historical documents and literature alike, no comprehensive review of the insecurities of hotline systems has been conducted since the 1990s. This paper concludes that there are many technical and political updates that can be made to improve the effectiveness of 21st century hotline communication systems in their mission to secure a path for peace in the face of nuclear war.
6. Regenerating Topsoil

Tule Horton | Full-Time Fellow
Mentor: Dr. Julia Novy-Hildesley, Stanford University
→ Learn More
Despite relatively little appreciation, topsoil plays a critical role in many of the climate change battles fought today - clean air, clean water, carbon sequestration, and food security. Yet topsoil degradation is occurring at record rates across the globe, largely driven by intensive agriculture and increasingly harsh weather conditions. Today, a third of the world’s soil is moderately to highly degraded and we continue to lose on average 24 billion tonnes of topsoil annually, a significantly larger amount than regenerates on its own. Topsoil provides the fertility upon which all other life thrives, and in this way, desertification presents a massive existential risk that threatens the prosperity and health of future generations.
Farmers are rarely recognized as frontline workers in the fight against climate change and existential risk and are more often denounced as culprits. Yet in fact, farm owners have perhaps the most at stake as well as the most potential to manage topsoil loss. Through regenerative agriculture practices such as no-till, cover cropping, and cultivating biodiversity, as well as the use of new technologies including biochar and precision agriculture, farmers are positioned to regenerate topsoil and restore fertility and abundance to our ecosystems. Yet political, economic, technological, and cultural barriers stand in farmers’ ways. In my research article I outline these challenges as well as preliminary work to overcome such hurdles and encourage the regenerative that will restore soil health and the human health that is often unknowingly but intrinsically linked.
7. How Much Do We Need to Know: A Case Study of Pandemic Influenza Risk in the United States

Jonathan Lipman | Full-Time Fellow
Mentor: Dr. Mehran Sahami, Stanford University
→ Learn More
Knowledge provides the power to invest appropriately to further our understanding of long term risks. This paper uses computer modeling (Bayesian networks) to help understand the risks from pandemic influenza and the assumptions estimates of those risks are based on.

We use sensitivity analysis to determine how greater uncertainty in selected model priors influences projected deaths, decimation, and mortality loss forecasts for pandemic influenza in the United States. We determined that the level of certainty of the probability of a pandemic emerging in a given year is the most significant factor for forecasting the chances that there will be a decimation of the US population over the next 100 years. We can use these insights to guide our prioritization of profiling other, less studied diseases.
8. Empowering the Vast and Voiceless Future: Lessons from Abolition, Democratization, Climate Politics, and Asilomar

Mauricio Baker | Full-Time Fellow
Mentor: Dr. Jeremy Weinsten, Stanford University
→ Learn More
Moral philosophers have compellingly argued that preserving the potential value of Earth’s long-term future is of immense importance. For such efforts to succeed, it may be crucial that the interests of future generations receive durable, substantial political protections, which are currently lacking.
What efforts would best contribute to such protections? To help answer this question, I first review insights from relevant fields, with a focus on decision-making psychology and institutional theory. Then, I investigate several historical developments that at least initially seem to be examples of voiceless groups receiving protection or empowerment, since we may want to do something similar for voiceless future generations: the abolition of slavery, expansions of the voting franchise, climate change mitigation, and genetic engineering governance. Drawing on these studies, I create and argue for a qualitative, rational-choice model that makes predictions about when shifts toward greater political inclusion occur and persist. I discuss the model’s implications for theory of moral circle expansion, political strategies, and institutional designs, with a focus on how people today can support future generations.

Zixian (Sunnie) Ma | Part-Time Fellow
Mentor: Jeffrey Ding, Future of Humanity Institute
→ Learn More
There is a Chinese phrase called 未雨绸缪, which means closing doors and windows before it rains. It is widely used as a commendatory metaphor to approve people of being prepared for a potential catastrophe. As closing windows before heavy rains prevents damages more effectively than when or after it rains, when it comes to artificial intelligence (AI), I believe it is also valuable to take preemptive measures to mitigate if not prevent risks from potential shock points in AI development and deployment.

A high-risk AI shock point, in the scope of this project, is defined to be a development or deployment in the field of AI that could cause unexpected and significant risks on our human society, thereby pushing some moral or ethical boundary, creating panic and, often, calls for heightened governance. Powerful yet potentially unsafe or unethical AI applications that count as high-risk shock points, such as autonomous vehicles being deployed prematurely, could cause irreversible and disastrous risks on human society, which could severely damage if not destroy human potential.
In this project, I propose an action plan named 未雨绸缪, which recommends actions on mitigating risks of advanced AI applications. Specifically, this project focuses on addressing emerging risks in China’s AI development and deployment as I am personally interested in China’s AI policy, have more knowledge and leverage in this space as a Chinese citizen, and expect an AI shock point to occur in China given a recent biotech shock point in China (i.e. the CRISPR’d babies experiment). It is worth noting that, although my focus is on mitigating risks of AI shock points originating from China, the scope of the risks concerned is still global, and many recommended actions can be adapted to apply in other countries’ AI development and deployment.
10. If Humanity Were One Human

Odelia Lorch | Full-Time Fellow
Mentor: Dr. Andrew Critch, Center for Human Compatible AI, UC Berkeley
→ Learn More
The neglect of the risk that humanity could destroy itself in governmental, academic, and social spheres is a shocking oversight. Yet even those who are committed to understanding and preventing forms of this risk struggle to fully grasp, both cognitively and emotionally, the enormity of these risks and their consequences. I set out to understand the cognitive biases that may be preventing people from grasping GCRs and exercising effective judgement, and to explore the potential of data visualization to counteract them and communicate compelling GCR-scale arguments.
Two of the most relevant cognitive biases I learned about were scope neglect, which prevents people from comprehending risks at a very large scale and proportionately grasping their magnitude, and optimism bias, which makes people more likely to live with a very low-probability but incredibly high-risk loss (such as a GCR) than they are to opt for an almost-definite small loss (such as concessions to prevent GCRs). While reading up on worst-case scenario risk communication, I learned one great way to counteract optimism bias is to evoke outrage. This makes people put more weight on the magnitude of the risk than its probability, and can be effectively done through emotional appeal, placing emphasis on the humanity and personal significance of an issue.

These observations, along with the image of humanity in its adolescence, the time period in which we have the advanced capabilities to destroy ourselves and have not yet developed the maturity to prevent such disaster, led me to the thought experiment that became my main deliverable.
We’re each aware of our own mortality, and grapple with this awareness throughout our lives. Yet as much as we may struggle to grasp our personal fates, it’s even harder to grasp the mortality of humanity. This project is a multi-page web visualization that prompts viewers to think about the future of humanity by asking the question, “what if humanity were one human?”