Training for Good

Red Teaming Topics

Red Team Challenge

This document contains a list of potential red teaming topics for participants in *Red Team Challenge*.

This list is by no means definitive and each team is free to choose any topic for their red team (whether it is listed below or not). However, we expect the majority of teams will find it most convenient to select a red teaming topic from this list.

There will be some time to discuss topic selection with your team during the workshop on Saturday so we advise you to give this some thought ahead of then. **Please let us know which topic your team has chosen by emailing cillian@trainingforgood.com by Monday 9th May at the latest.**

Philosophical & intellectual foundations

- The 21st century could be the most important century ever for humanity. See Holden Karnofsky's "Most Important Century" blog post series. (Note: you may wish to choose **one** of the key claims / articles in this series to focus on for your critique).

- The expected value of extinction risk reduction is positive. See The expected value of extinction risk reduction is positive and this response A longtermist critique of "The expected value of extinction risk reduction is positive".

- The arguments and conclusions outlined in this Open Philanthropy post: Worldview Diversification.

- That hits-based giving is the best approach for large EA donors, such as Open Philanthropy, to take. See: Hits-based giving.

- The arguments and conclusions outlined in this GPI paper: The epistemic challenge to longtermism

- That aspiring EAs should aim to maximise the expected value of their actions. See: Why Maximize Expected Value? (Note that you may wish to find other relevant sources which lay out the reasoning in greater depth).

AI Safety

- Open Philanthropy's reasoning for considering potential risks from advanced artificial intelligence a major priority. See original post from 2016: Potential Risks from Advanced Artificial Intelligence: The Philanthropic Opportunity.

- The arguments and conclusions outlined in this EA Forum post: AI Governance: Opportunity and Theory of Impact.

- The arguments and conclusions outlined in this Cold Takes post: AI Timelines: Where the Arguments, and the "Experts," Stand

- The arguments and conclusions made in any one of the following reports commissioned by Open Philanthropy:
    - How Much Computational Power It Takes to Match the Human Brain
    - Forecasting Transformative Artificial Intelligence with biological anchors (draft report by Ajeya Cotra)
    - Report on Semi-informative Priors

Global Health & Development

- The arguments and conclusions outlined in this EA Forum post: Growth and the case against randomista development.

- Snakebite treatments may reach effectiveness in the same order of magnitude as Givewell's recommended charities. See Snakebites kill 100,000 people every year, here's what you should know.

- The arguments and conclusions outlined in this Rethink Priorities report: Intervention report: charter cities.

- The arguments made by GiveWell for believing that any one of the following organisations should be considered a "Top Charity".
    - Malaria Consortium
    - Against Malaria Foundation
    - Helen Keller International's Vitamin A Supplementation Program
    - New Incentives
    - Evidence Action's Deworm the World Initiative
    - SCI Foundation
    - The END Fund's Deworming Program
    - Sightsavers' Deworming Program
    - GiveDirectly

- The approach of the Global Health and Development Fund. Specifically, you might consider arguing that:

    - The Global Health and Development Fund does not improve the health or economic empowerment of people around the world as effectively as possible.
    - The fund should not be managed by GiveWell co-founder Elie Hassenfeld and other GiveWell staff.
    - The fund should accept funding applications (as the Animal Welfare Fund, the Long-Term Future Fund, and the EA Infrastructure Fund do).

Animal Advocacy

- The arguments made by Animal Charity Evaluators for believing that any one of the following organisations should be considered a “Top Charity”.
    - Faunalytics
    - The Humane League
    - Wild Animal Initiative

- Wild animals have net negative welfare (i.e. wild animal suffering dominates wild animal happiness). See Center on Long-Term Risk’s The Importance of Wild-Animal Suffering.

- The arguments and conclusions outlined in this Rethink Priorities report: Invertebrate welfare cause profile

Longtermism

- Preventing global catastrophic biological risks is one of the most pressing problems. Specifically, 80,000 Hours’s reasoning for listing “reducing GCBRs” as one of its highest priority areas. See: Reducing global catastrophic biological risks.

- Choose any one of the specific project ideas listed by the Future Fund. Assess the arguments for believing that this project could be among the top projects for positively impacting the long term. Future Fund Project Ideas.

- Choose any one of the biosecurity projects listed in Concrete Biosecurity Projects (some of which could be big). Assess the arguments for believing that this project “could reduce catastrophic biorisk by more than 1% or so on the current margin”.

- Choose any single chapter in *The Precipice*. Assess the arguments and conclusions of that chapter.

- The arguments and conclusions outlined in this EA Forum post: Should we buy coal mines?

- The arguments and conclusions outlined in favour of patient philanthropy. See: GPI Report: Patient Philanthropy in an Impatient World, 80,000 Hours Podcast: How becoming a ‘patient philanthropist could allow to do far more good or EAG Talk: Philanthropy timing and the hinge of history.

Policy and party politics

- Choose any one of the following Charity Entrepreneurship research reports on health and development policy. Assess the arguments for considering this an idea worth recommending to future charity founders:
    - Tobacco Taxation
    - Aid Quality Advocacy
    - Road Traffic Safety

- Aspiring EAs should pursue AI policy careers in the EU. See EA Forum posts on this topic: Should you work in the European Union to do AGI governance? and Argument Against Impact: EU Is Not an AI Superpower.

- That the EA movement should pursue policy as a path to impact as argued in this talk: Should EAs do policy?. Specifically, you might consider arguing in favour of one or more of the following claims:
    - It is hard to pick good policies
    - Policy change is intractable
    - Policy-makers are already proto-EA
    - Politics could hurt the EA brand

- The arguments and conclusions outlined in this 80,000 Hours post: If you care about social impact, why is voting important?

- The arguments and conclusions outlined in this GPI report: Longtermist institutional reform

EA community building / meta charity efforts

- Choose any of the views on EA movement growth presented here. Assess the arguments for believing this view to be incorrect or damaging in some way. (Note that you may wish to find other relevant sources which lay out the reasoning in greater depth).
    - Effective altruism should be broad
    - It should stay narrow
    - It should focus on having a great culture first
    - It should take a higher variance approach to recruitment
    - We should be wary of committing to one direction
    - We should consider people who aren’t ready for a big commitment
    - We should reach out to influential people
    - We should want more exposure for both good and bad ideas

- The arguments and conclusions outlined in this EA Forum post: EA needs consultancies.

- That aspiring EAs can greatly increase their impact by building the right career capital and that career capital should likely be one of the top considerations early in their career. See Career capital:how best to invest in yourself.

- An EA Forum contest for critiques and red teaming is a good idea. See: Pre-announcing a contest for critiques and red teaming

- CEA's decision to discontinue its focus university programming. See: CEA is discontinuing its focus university programming, passing funding to Open Philanthropy AND/OR CEA's decision to launch the Campus Specialist programme in the first place. See: A huge opportunity for impact: movement building at top universities)

- The arguments in favour of a "common application for Effective Altruism" as outlined in this EA Forum post: Brief Presentation and Considerations for an EA Common Application

Miscellaneous

- Choose one of the career reviews listed on the 80,000 Hours page The highest-impact career paths our research has identified so far. Assess the reasons for recommending this career path and scrutinise how this review might turn out to be unhelpful, misleading or counterproductive to an aspiring EA.

- The arguments for considering space governance to be among the highest priority areas. See: Space governance.

- Personal assistants make people meaningfully more productive AND/OR aspiring EAs should consider working as a personal assistant to someone in a high impact role as a means of themselves having an impact. See: To PA or not to PA?

- There are no obvious ways that [EA organisation's*] strategy could be better optimised towards achieving that organisation's stated goals and/or the broader goals of doing the most good.
  - *E.g. CEA, Open Philanthropy, GiveWell, FTX Future Fund, Charity Entrepreneurship, HLI, etc.