

Organise the next Virtual AI Safety Unconference

Summary

I have a design for an online unconference, that I have run a few times. It's tried and tested. But I don't have the time and energy to keep running them anymore. I would like to find two people to take on the task of running the next Virtual AI Safety Unconference (VAISU). Even though I have a ready format, there is room for you to improve the event design too.

The goal of this project is both to produce the event, and also to pass on my organising skills to people who will hopefully use them in the future. I'm therefore looking for team members who are interested in continuing on the path of being organisers, even after this project.

I'll be acting more like a mentor than a full team member. I'll teach you as much as I can, but you will do all the work. The reason I'm proposing this project is because I don't want to organise the next VAISU, I want you to do it.

The non-summary

Motivation

Online unconferences are low cost and relatively low effort to run, and are also really great. I think this type of event has the highest cost to value ratio out of any event type that I know of.

Why more events?

AI safety is growing faster than the infrastructure can keep up with. Every AI safety event I know about gets a massive number of applications. There is definitely room for more events.

Why unconference?

Unconferences are great in many ways, but they are especially great in a rapidly growing movement, where the people at the top, can't keep up with or even keep track of the need for infrastructure. Unconferences are self organised around the needs of interests of the participants. This makes them more adaptive but also lower effort to run relatively to other types of events.

I'm not saying that unconferences are a one-size-fits-all type of event. There is definitely room for more specialised events too. But everything equal, I think AI safety would be better off with more unconferences.

Why online?

Online and offline have different pros and cons.

One big advantage with online events is that they are much more accessible. This is important! Even if we have lots of in person events, it's worth having online events too. Accessibility helps in two ways. Firstly there are people who can't attend otherwise for various reasons. Just paying for their travel may not help since the barrier might be time or VISA. Secondly, it lowers the cost (in time and money) to attend for everyone. People who would not be willing to pay the time cost for an entire in-person event can choose to join in for just a few sessions of an online event, benefiting both them and everyone else (I've done this myself). Allowing people to pick and choose among sessions, rather than commit to a full event, opens up for more specialised sessions.

The other main advantage of online events is that they are almost infinitely scalable, much less work to organise, and much lower organising costs. A single person can do 90% of the work running an online unconference. There are some things that need to be done in parallel during the event (requiring multiple people), but I have never had a problem getting enough volunteers to help out with that.

Project plan

First step: Design the event

What tools are we using, why and how? How will it all come together?

As mentioned, I already have a recipe that I've used for previous events. But my experience for previous attempts, is that I can't just hand over how to do things, and expect it to work. In order to understand the event design, you have to be part of the design process. This means you will re-evaluate the design decisions I've made in the past, and possibly come to different conclusions for how you want to do things.

If you decide on a design which I think is worse than I would have done, I will argue with you, but I will not overrule your choices.

Second: Announce the event

You want to announce the event as early as possible (ideally a few months in advance for this size of event) so that participants can find out about it in good time.

By "Announce the event" I mean anything you need to do in order to announce the event, and nothing else. This includes:

- Deciding the date
- Decide where and how to post it. Do you prepare a website already, or just a "save the date post" on LW and EA Forum? Or something else?
- Writing an announcement draft
- Imagine a potential participant reading the announcement.
 - Notice what this imaginary person wants to know.
 - Notice what decisions you need to make now, in order to tell this person what they need to know.
 - Make those decisions.
- Finish writing the announcement and post it.

Next steps: Everything else

I don't have a list of everything that needs doing. The way I run events is that I run a simulation of the event in my head, and this simulation will tell me what needs doing. I tried to write a list of all the different event ingredients [here](#), but I've probably missed something important.

A lot of the work is just paying attention to everything and notice what needs doing.

Required reading

These posts are not written by me, but fits in very well with how I do things. If you like these posts and vibe with this way of doing things, then you're probably a good fit for this project.

- [How To: A Workshop \(or anything\) — by Duncan Sabien](#) (The first part. Only read the "The Rest Is Commentary" section if this seems useful for you.)
- [Community building: Lessons from ten years of facilitation experience — by Severin](#)

Who should and shouldn't join this project?

This is an organising project. Joining this project will put you on the path to becoming an organiser. You will not be stuck in this path. It's possible to move back and forth between organising and research, as I have done. But you should also know that my organising work has slowed down my research career (by taking up a lot of time) more than supporting it. I have no regrets, but I also want you to know what you will and will not get out of this project. This project is not a stepping stone on your path to become a researcher, i.e. it doesn't bring you closer to having a research job at a research org.

This project will let you have a direct impact by doing this event, and you will learn online organising skills. If this sounds good to you, then you are welcome to apply. I think it's possible to have a massive impact by running great events, and you'll get a lot of appreciation from attendees.

I'm especially interested in team members who want to continue on the path of being an AI safety organisers, and run more events like this one.

Some previous Virtual AI Safety Unconferences

I've run (more or less) this event 3 times now, with different collaborators and changing the name slightly every time.

- [Web-TAISU 2020](#)
- [AISU 2021](#) with Aaron M. Roth
- [VAISU 2023](#) with Nguyen Tran

There have also been some other similar events that I have not been involved in.

Output

The main output will be to organise the event itself.

An additional output could be to produce some sort of write up, but this is less important.

Risks and downsides

More and more AI safety projects come with risks of accelerating AI capabilities. This is bad, since we don't want AIs to become too capable before we know how to make them safe. Or if alignment is not possible, we want to get clarity about this before it's too late to stop AI progress, and accelerating AI capabilities reduces the amount of time we have left.

This means that creating tools and infrastructure for AI safety research (e.g. this event), also comes with the risk of increasing the dangour. So far, in my experience, with the events I've been involved in, I judge it to be clearly net positive. However, this is something any AI safety organiser needs to keep an eye on. The field is growing and changing. This means that the types of people that join an event might shift from year to year, changing the balance of risk and benefits.

Acknowledgements

Thanks to my previous co-organisers, Aaron M. Roth (AISU 2021) and Nguyen Tran (VAISU 2023).

Thanks to [EA Hotel](#) for housing and feeding me for most of 2019, and giving me the time and opportunity to explore many things, including some experimental events, which put me on the path I am now.

Team

Team size

Probably 2 ppl + me giving advice/mentorship.

If I get a lot of great applicants, I might expand both the team size and project ambition in some way.

Research Lead

Linda Linsefors

linda.linsefors@gmail.com

I've been focused on AI safety in various ways since late 2016. Since then I've done a mix of learning, organising and research. For more details see [my resume](#)

My intention is to spend ~5h per week on this project. I'll be acting more like a mentor than a full team member. I'll teach you as much as I can, but you do all the work. The reason I'm proposing this project is because I don't want to organise the next VAISU, I want you to do it.

Team Coordinator

Even though the team is very small I think it would be good to have a team coordinator who is in charge of planning and running meetings with the three of us. I would like to not do this, which means that this team is looking for a team coordinator.

Skill requirements

Required:

- Interest in event design.
- Good at extracting information from me, and noticing what you don't know yet.
 - I have not been very successful at delegating in the past. Key Information keeps being lost in transmission. I'm therefore looking for team members who will keep asking questions until you actually have received all the information you need.
- Good written communication.
 - When you produce some written information for the unconference participants, you need to be able to look at your own text and tell if it's good enough, and improve what needs improving.
- Paying attention to *everything* and notice what needs to be fixed or improved.
 - You don't necessarily need to know how to solve a problem. Just noticing that there is a problem and asking me for advice is good enough.
- Willingness to do some amount of boring spreadsheet tasks.
 - You're welcome to try to automate as much as possible, but I predict that you will not be able to automate away everything.

Non essential but very useful:

- Previous organising experience
- General knowledge about AI safety
- Know how to set up and use various online tools
 - Discord
 - TeamUp
 - Coda
 - Google sheet
 - Gatheround
 - Etc
- Programming skills
 - I've written a python script for scheduling. Someone who is better at programming than me could probably improve.
- Public speaking skills (for opening and closing sessions)
- Knowledge and/or connections for getting sponsorship for this event