The AI alignment community has historically been very cautious about outreach (possibly too much so) because of worries about creating a bad impression of AI safety concerns and bringing in badly aimed mass attention. Low quality outreach can indeed be net harmful, and some people really shouldn't be working on this, but for some people it seems like a reasonable choice, especially now that the "cat is out of the bag" and the wider public is becoming more aware of the possibilities and risks of AI. Further outreach won't give people their first impression of the AGI safety project, but it can replace their impression with a more informed (or more misinformed) one.

Rob Miles uses a standard for public outreach where, if he puts a video out, it has to be the best video on a particular topic. (It can be the best in some specific way, or for some specific audience.) You can use a somewhat lower standard for things like podcasts and assume it'll be worth it as long as the podcast is quite good, and use a lower standard still for things like talks, where it doesn't matter if it's the best.

If you're doing this kind of work, it's important to have a strong understanding of the issues. Keep reading relevant materials and be familiar with the questions that people ask most often.

To find out if you're good at it, and to get better, start with low-profile projects and ask trusted sources for feedback on how you could improve. If it seems to be going well, you can repeat this with increasingly high-profile projects over time.

One area that could have a significant impact would be the creation of materials about AI safety in non-English languages like Chinese or Russian.

Related

- How can I work on AGI safety outreach in academia and among experts?

Scratchpad

From old doc

- Rob Miles principle: If he puts a video out, it has to be the BEST video on a particular topic. That's a great standard for mass outreach.
- The seeds are already out there, most relevant people have heard of the AGI safety project. Mostly, you can only replace the seeds by better or worse ones, not just get people on board from scratch!
- Discourage bad outreach
- Emphasize importance of local knowledge: Maybe AIS materials in Russian, Chinese, … make sense? We don't know!

plex: “similar to 2.15, generally understand the stuff; get feedback; appear in places where you can get feedback from people who will give honest opinions, only continue if you get good feedback”
“There has been a norm against outreach, possibly overly strong; some people shouldn’t (not all outreach is good), but pretty reasonable that some people should, especially now that the cat is out of the bag”
“RM principle is good for video content but doesn’t quite apply to podcasts; just need pretty darn good podcast, and for talks it’s different, doesn’t need to be the best talk”