

Assessing the Potential of Congressional Messaging Campaigns for AIS

Summary

This project aims to figure out if congressional messaging campaigns (CMCs) work, and if they do, what messages of AI concern to promote, and how to promote them in a high-quality manner. At minimum that will involve researching general CMC effectiveness and writing a report, but it will likely extend to leveraging this research to develop a best strategy for deploying a CMC for AIS. Time permitting, we'll take our findings and deploy that best strategy, attempting to help fill the void of actionable steps on AI risk for those less involved.

The Project

Theory of change

We don't have polls on AI going back very far, but we have at least [one per year](#) since 2017, the majority taking place in the last year, which should give you some idea of the recency of this issue and how quickly it permeated the political consciousness. The most direct finding to highlight over time is that of concern about AI, growing from [39%](#) in 2018 to [68%](#) in 2023. But more than concern there's also support for regulation: [62%](#) support creating a federal agency to regulate AI, [65%](#) support required safety testing, and [24%](#) support one of the most radical solutions proposed (a ban on all new research). So this is my first claim: the majority of the public is concerned with AI, and that there is support for regulation.

Things move slowly in Washington, and there are multiple potential proposals in the works right now for regulating AI. So I can't point to the lack of a passed bill right now and say "look Congress isn't doing *anything* about AI" and some might even go on to say "we already seem to have what you want, bills in progress, why is that not enough?". I think my basic answer is that sentiments in Congress don't reflect the public at large right now. The concern driving the bills seems to flow largely from a few concerned congressmen, and though there are certainly some congressmen who disagree and are more ready to embrace developments in AI, most just don't seem to be all that occupied with it, more concerned with other matters. Thus my second claim: public concern on AI does not match concern from lawmakers.

But we're still missing one key ingredient. In a representative democracy, the theory is that representatives make decisions based on public opinion, but to what degree is this true? [Burstein](#) (2003) found that "the impact of public opinion [on policy] is substantial" and that responsiveness seems to be stable over time. Some protest that it's really money that

drives policy, but there's work to indicate against this, [Ansolabehere](#) (2002) finding contributions have very little effect on legislative voting. Others protest that it's powerful elites and political organizations that control policy, but [Burststein](#) found that public opinion is impactful even after taking influence from political organizations and elites into account. We can also look to another emerging issue, climate change, and note that this is an issue that has been responsive to changes in public opinion, [Anderson](#) (2017) finding public opinion to have a "significant and positive effect" on climate policy across Europe. What's more is that concern for AI might be greater than that for climate change, a [2018 survey](#) finding 29% "very" and 40% "somewhat" concerned about climate change, compared to a [2023 survey](#) that found 43% "extremely" or "very" and 33% "somewhat" concerned about risk from AI. This is my third claim: public opinion matters for policy.

So public opinion matters for policy, but how much change comes from communication of public opinion? Or more specifically, how much comes from CMCs? A [2018 survey](#) from the Congressional Management Foundation provides a promising finding that as few as 50 emails are enough to influence the legislative agenda. But further backing is hard to find quickly, as the research (on a first glance) is sparser (hence part of the goal of the project). If nothing else, I think the above finding supports my fourth claim: looking into CMCs further could be beneficial (and I think it will likely be an effective means of influencing policy).

The final question we might ask is: what is the current AI safety space doing to make use of this situation? Currently I would argue very little, at least very little optimally. There's nothing like this for the US. There is a [limited version of this](#) for those in Europe, but this doesn't seem to have had wide deployment and also has made some assumptions that might not be great (i.e. that Word is a good format to use, when many don't have access to Microsoft Office). And beyond being an opportunity that we could take advantage of, I think this could enhance some of the work those in AIS are already doing. Interventions aimed at capturing the attention of the public (e.g. [Rational Animations](#)) don't have a current call to action for those concerned, but without the desire or time to significantly involve themselves. So my fifth claim: there is an open opportunity here for AIS to leverage public concern, as well as add to other methods of outreach.

So what do we have when we add it all up? The public is concerned with AI. Their concern isn't matched by lawmakers. Public opinion has a substantial impact on policy. CMCs are one way to get their attention, raise awareness, and (potentially) lead to change. AIS could make use of something like this. Hence, this project.

Potential Schedule

The **least ambitious version** of the project involves at least outputting a research report on the effectiveness of CMCs if we find the research indicates they are likely not effective. The **moderately ambitious version** of the project, involves tailoring a CMC to AIS, red teaming that, and then incorporating feedback to create a forum post on the most promising CMC for AIS. The **most ambitious version** of the project will involve taking that best method and deploying it.

Phase	Goal	Work
-------	------	------

Plan (1 week)	Get to know each other and create a plan moving forward (discuss this table and update it as we see fit).	Reading this document in total and preparing any changes you think we should make for a meeting together.
Research on the effectiveness of CMCs (3 weeks)	Research whether MCs can lead to increased legislative change and develop a shortlist of the best strategies. Also decide whether there's sufficient cause to continue the project	Individual research for the first two weeks, as well as outreach to EA orgs and others in the policy space to get their views, with a meeting at the end of the third to come to conclusions
Writeup on effectiveness of CMCs (1 week)	Collate all the research into a well written write-up on the effectiveness of CMCs, and post it.	Everyone contributing to a google doc and helping prepare it for posting by the end of the week
End Least Ambitious Version		
Research characteristics of those concerned with AI risk and choose strategy (1 week)	Develop a better understanding of those concerned with AI risk and use those characteristics to choose the best strategy for AIS.	Individual research as well as fleshing out our current best proposal for a CMC for AIS
Red Team our best idea (1 week)	Create a list of reasons why our best strategy could go wrong, and invite comment from others	Reaching out to people with experience and general internal red teaming
Solidify our best method for a CMC for AIS (and write a forum post) (1 week)	Incorporate red teaming feedback and create a comprehensive overview of what and how we would deploy a CMC for AIS and translate this into a forum post if we decide to end here.	Editing the proposal in the google doc and meeting at the end of the week to determine if we want (or have the time) to continue forward or not.
End Moderately Ambitious Version		
Develop output and start outreach (3 weeks)	Develop our chosen deployment (likely a website) and outreach to everyone in the space we think it might be useful for.	This is where work might diverge a bit depending on skill set, where someone with web dev experience would develop full time while others focus on outreach.
Seek feedback from others on deployment output and fix (1 week)	Making sure the deployment product is working smoothly and finalize it.	Get people to test out the deployment and incorporate feedback for any errors.
Deploy and plan for future (1 week)	Take the deployment method (likely a website) live and encourage people to update with links to it and figure out future plans for the project.	Encouragement to start linking the project, and planning for the future (e.g. answering how we'll keep the website maintained).

Risks and Downsides

- **Crying wolf:** pushing for congressional action and change now could reduce capacity to do the same "in n years when things are really critical, but people have become inured to it"
 - Mitigation: explain the potential harms of AI, but leave the timeline open for when harms could materialize. But also there's a:

- Difference of opinion: this is sort of a catch all for doing any AIS work that is focused on current capacity building and action instead of setting up capacity for further in the future, and while my timelines aren't crazy short, they have wide bounds that extend into the medium term future.
- **Political polarization:** though there's nothing inherently partisan to communication campaigns, any intervention in the space that promotes engagement in the political sphere could lead to polarization and as such should be addressed.
 - Mitigation: it will be made clear throughout that this is an issue of bipartisan concern (there's also a fair bit of polling to back this up), and that there are many routes to regulation that both parties can support ([Blumenthal-Hawley bill](#) is a good example here) that we can steer things away from becoming partisan.
 - Mitigation: I'm also working on a current report on how issues in the past became partisan, so this is an issue I'm aware of and actively learning more on, which doesn't mean I have the answers but does mean I have at least some guidance.
- **Time waste:** I could be quite bad at managing the project fully and cause sup-par allocation of time and talent, thus leading to a worse use of time than other alternative participants might have
 - Mitigation: I hope to check in somewhat regularly and have a fairly democratic process that allows the team to go in the direction everyone thinks as best
 - Mitigation: I also would do some reading up on how to manage others, and would make sure that I'd at least got some background material under my belt so that I could be a better manager than I am now before the start of the project.

The Team

Members & Positions

- **Targeted Team Size:** 3-6 people
- **Research Lead:** Tristan Williams (Whatsapp: +12767062529)
 - **Background:** AI Safety RA, Contemplative Sciences RA
 - **Timezone:** UTC +1
 - **Time Dedicated:** I expect to contribute roughly 10h/week, but can contribute up to 20h/week
- **Team Coordinator:** Happy to do it myself, but also happy for someone to take on this role

Skills/Experience Needed:

- **Absolute:**
 - **Policy Experience:** Whether in research or something like working in a congressional office, this project needs additional expertise in this area. I have some knowledge of the area having worked as an RA in the policy space for a year now, but we will need more experience if we encounter little

information on MSCs, so without someone that fits this bill the project will not continue. Generally, I recognize that I'm bringing less experience to the table here than some other projects, and as such am quite okay with a flat hierarchy and will be happy to onboard those more experienced and defer to their judgment often.

- **Base:**
 - **Generalist research skills**
- **Preferred:**
 - **Communications:** we'll likely be doing outreach to a number of orgs to try to get an idea of how their needs interact with our best strategy and to facilitate deployment of the final product, as well as policy people to red team our idea
 - **Writing Well:** to make our report accessible and allow others to pick up specific actions where we left off
- **Super Desirable:**
 - **CMC experience:** this would be golden. And to a lesser extent, experience with grassroots organizing generally would be helpful in the same vein
 - **Web design:** in case we want to test a messaging approach
 - **Policy making:** it would be great to have some insight into the policy making process and specifically how congressional offices work so that we don't have to build all this information from the ground up
- **What We Already Have:**
 - **Research:** My experience is mostly research oriented, so I'll likely be most helpful when it comes to figuring out the "what should we do?" part of this.