

## Statements on AI Risk: Using Public Moments to Advance AI Safety

**Sharing rules for this document:** Don't publish this document. Do not copy all or parts of this document into other published documents. This document can be shared in whole with organizations, institutions, and individuals that work on reducing risks from artificial intelligence.

### Motivation

There have been several recent public statements on AI risks. These include the publication of the [open letter by the Center for AI Safety](#), the [Pause Giant AI Experiments Open Letter](#), and [Eliezer Yudkowsky's Time op-ed](#). For proponents of AI safety, understanding the theory and application of social persuasion can help make the most of these statements. While public moments in AI are relatively new, history is filled with similar sparks of change.

In particular, spikes in public interest are not only public relations opportunities, but also vital political- and policy-building openings.<sup>1</sup> The increase in interest means some people in influential positions are more reachable and have greater attention to the issue (for now).

We have put together this brief guide in the hope that it will aid you. Some elements of it will hopefully seem obvious, but others may be new. If you want to go beyond this guide, see the last section.

### Three immediate actions:

- Invest in improving your positioning in political spheres.
- Use social science tools to empower your influence.
- Focus on big ideas more than policy details.

---

<sup>1</sup> Here, "politics" and "policy" means not just human decision-making in governments. It includes other social structures that could influence AI safety. For example: companies, coalitions and more.

## Invest in improving your positioning in political spheres

*Find the few that already agree, help them be better insider advocates*

Advocacy is often understood as an activity of persuading or pressuring decision makers into action. Typically, a far more effective strategy is to find a few internal supporters of your cause and leverage their existing networks. Information flows better and faster through established networks of trust.

In this approach, your role shifts from a communicator to doing the “legwork” in the background, briefing these internal supporters on key messages and facts. Be ready to support them with any information or capacity they might need to be an effective advocate within their walls, and at other institutions where they might also wield influence.

Analyze whether you already have internal allies. If you do, ask what support they need and offer ideas for support you can provide. If you don’t, think about places where potential allies might gather (online and offline); go there and discretely offer to share your subject matter expertise.

*Install yourself as a policy advisor with key decision-makers by making yourself a source of support*

1. Analyze the stakeholders with whom you currently lack an established relationship. Which of these will significantly influence the governance of this issue?
2. Reach out to the highest level policy-maker you might get a response from. It is usually much better to get delegated down than to aim too low. Reach out and offer to help them respond to this immediate development, which you are able to do because of [insert your capabilities].
3. Once you have a meeting confirmed, gather information:
  - a. Contact relevant junior staff of institutions/policy-makers, expressing your willingness to provide information on AI risks. Request a conversation to address expertise needs and pressing questions.
  - b. Aim to identify a problem that the policy-maker has (this might just be "getting greater clarity on the issue") and demonstrate that you can be helpful in solving this problem in your meeting.
4. Strategically prepare a communication strategy with key messages, using the tools below. In the conversation, aim to also demonstrate expertise in the recurring

challenges that surround this public moment, not just in dealing with the moment itself. Establish yourself as a source of information they can return to again and again.

## **Use social science tools to empower your influence**

### *Offer an information advantage*

“Many are thinking at the moment about the increasing risks from AI, but not that many know that [tease the unique information advantage you can offer]. If it is of interest to you, you could be among the first to get a comprehensive overview about [your information advantage].”

### *Use dynamic social norms*

Highlight how this topic is gaining traction, but that it is still possible to be among those who were early to the issue. For example: “In the last weeks, more and more [of your peer group] have realized the dangers of unchecked AI development.”

### *Make it as high-status as possible to care about AI risks*

For example: “The brightest, most-well respected scholars in the field of computer science, and the industry leaders in AI share this concern. As a responsible policy-maker with strategic foresight, we know that you understand the value of expert concern. We think you’d want to be/your country/company wants to be among the first who take more serious action on this threat.”

### *Build a golden bridge*

Make them feel good about changing their minds, instead of making them feel bad about the fact that they haven’t done that already: “Until recently, it was very difficult to get greater clarity as a policy-maker about how grave the risks of AI are. Now that there is greater clarity, from the communications by scholars and industry leaders, we appreciate that you take this seriously and are trying to inform yourself about the challenge.”

### *Affirm a developing identity*

Make statements about who they are that are easy for them to agree with, but hard to disagree with: “[The fact that you are meeting with us/that you already said XYZ in public] clearly shows you are among those who understand how dangerous unchecked AI could be.”

### *Use a ‘control’ frame*

Avoid a “powerful AI” frame, at least as an isolated idea. For many, having something powerful is quite positive. Instead, use (or add) the notion of humanity developing out-of-control AI. Control is one of the most fundamental human needs. Avoiding a loss of control is recognized as a valuable goal across political ideologies. Even weak-but-out-of-control AI is not appealing, let alone a powerful-but-out-of-control one.

### **Focus on big ideas more than policy details**

For example, talk about institution building rather than a specific regulation. Getting agreement on specific policies requires there to already be general agreement that something is a problem or an opportunity. Public moments are usually better at normalizing a general concern, less good at elevating wonky policy ideas. Jumping into details, without first building this shared concern, can often lead to people getting pulled apart prematurely by values and ideologies embedded in those details. However, if you can build a wave of momentum for high-level action first, it can carry people through the inevitable differences of opinion that will emerge in detailed debate.

See also our [How to Win Campaigns and Influence Policymakers: A Guide to Normalizing Ideas in AI](#).

### **What now?**

Need help taking advantage of this moment, or preparing for the next big milestone in AI? The tools above are just a sample of our expertise. Future Matters specializes in facilitating social and policy change. We draw on our team’s experience in policy, politics and movements, and a proprietary database of evidence-based social science interventions. From

multi-day workshops to rapid consultations on time-sensitive issues, we design custom strategy consultations to best fit your budget, time, and objectives.

For a free scoping discussion, please reach out to our AI Strategy Consultant, Kyle, at [k.gracey@futuremattersproject.org](mailto:k.gracey@futuremattersproject.org).