Sensitive - please only share upon permission

Diplomacy for US-China rapprochement on frontier AI

How to engage with this document

The memo below reflects a view our team deems plausible and would like to test and refine. We're particularly looking for input on the following questions:

- What parts of the arguments seem misled, wrong or unclear?
- What parts of the recommendations seem unrealistic, naive or ineffective?
- Where do you suspect SI could add most value?

Please have a low bar for reaching out for a 1:1 to discuss this or any related points.

Summary

The United States needs a two-pronged approach to manage China's AI capabilities. While export controls and other restrictive measures help counter authoritarianism, a lack of foreign service capacity is generating misunderstandings among rivals and allies alike. Especially in the context of frontier AI, a lack of mutual understanding is pushing China toward more desperate strategies, as well as alienating US allies and global majority countries. To reduce fear, doubt, and uncertainty, the US government's objectives would be well served by increasing its diplomatic capabilities. Diplomacy offers proven, cost-effective ways to reduce risks from drastic actions by contextualizing them using established norms and leveraging existing communication channels. Yet, American diplomacy has declined to below the levels of the 1950s. SI can help fill this gap by strengthening diplomatic channels and facilitating constructive US participation, leveraging our independent position and established relationships with governments across the globe.

1. A restrictive stance towards China requires careful diplomatic calibration

A restrictive stance can manifest in two distinct ways: through managed, strategic competition or through confrontational, all-or-nothing narratives. As many world leaders question either the ability or the genuine intent of the US to build a free world, absolutist narratives stand slim chances of success. The desire for America to decide humanity's future without regard for other approaches is reducing liberty's chances of success for several reasons.

1. China's strong ethnonationalism paired with industry overcapacity likely heightens the risk of escalation in response to US pressure, potentially pushing China toward a more aggressive posture, similar to Russia;
2. Lower and middle-income countries (LMICs) might align with China due to easier access to Chinese foreign aid and technology;
3. The continued advancement of open-source AI could enable capabilities diffusion even if hardware controls succeed; and
4. US allies, constrained by weak governments and mixed economic incentives, may be reluctant to join a confrontational approach, which increases costs for the US.

Therefore, while maintaining a restrictive stance is crucial to counter authoritarianism, it must be balanced with extensive diplomatic engagement to prevent escalation. Should China suspect the emergence of a novel decisive strategic advantage for the US, a nuclear first strike could be perceived as a last resort to counteract the liberal 'existential threat'. Governmental and philanthropic investments in diplomatic channels represent an effective strategy to foster mutual understanding on critical frontier AI issues.

2. Diplomatic channels: a neglected yet necessary opportunity

Investment in diplomatic channels for frontier AI serves three objectives:

1. Establishing crisis management mechanisms to de-escalate tensions
2. Developing robust AI guardrails based on shared understanding of its potential
3. Securing trust and support from LMICs by creating clear pathways for technological advancement (e.g. outline pathways from tier 3 to 2 to 1)

If diplomatic channels are layered sufficiently, they form a web of interconnections that make relationships between countries less likely to break down in case of tensions or instability. Diplomatic channels take three forms and tend to deliver significant value relative to their cost.

1. Track 1: Direct state-to-state engagement (institutionalized (UN, OECD) or not), for instance:
    a. UN peacekeeping operations: 2x more cost-effective than US military deployment
    b. Chemical Weapons Convention negotiations: $25M over 4 years, eliminated 98% of declared stockpiles
    c. International Civil Aviation Organization: $100M annual budget enables safe global air travel
2. Track 1.5: State engagement including non-state actors, for instance:
    a. Pugwash Conferences ($2M annually): Crucial role in Nuclear Non-Proliferation Treaty
    b. Geneva Initiative ($12M/5 years): Developed detailed Israeli-Palestinian peace proposals
3. Track 2: Non-state actors with government connections, for instance:
    a. Sant'Egidio Community mediation ($8M): Ended 16-year Mozambique civil war
    b. International Crisis Group ($20M annually): Early warning and policy solutions for conflict prevention

The choice of track depends on the political climate, the issue of concern and the sought outcome, but a combination of tracks usually leads to more robust relationships.

3. The state of diplomatic channels for frontier AI governance

While track 2 channels have increased and shown early results and promise for US-China cooperation, having primarily technical non-state actors talk is not enough. They may align on standards and guardrails, report back to their governments and secure specific actions, but this is unlikely to translate into lasting policy change across government. To propagate lasting changes, national administrations need to officially adopt taxonomies and commitments by institutionalizing

exchange between across all relevant departments. When it comes to interaction between administrations, recent progress includes:

1. Bilateral US-China talks
2. Signing of the Bletchley Declaration
3. The UN General Assembly adopted three AI documents in 2024, two resolutions tabled by the US and China respectively, and the Global Digital Compact, which mandates the creation of an International Scientific Panel on AI and a Global Dialogue on AI Governance.

On top of bilateral track 1, track 1.5 and 2, the US should focus more on institutionalized channels – such as the UN – because this is where authoritarian regimes prefer to engage as these fora recognize their sovereignty and aim for peaceful coexistence rather than imposing compromise. It is also where LMICs invest the vast majority of their scarce diplomatic capacity as the global majority.

The US has been increasingly reluctant to engage in institutionalized channels, despite having created them. After World War II, the US and its allies created a rules-based order that would export their priorities and worldviews globally – often in good faith. However, this system faced three major challenges: Russia failed to integrate into this rule-based order, US allies took the system's effectiveness for granted, and China's economic integration did not lead to the democratic reforms Western countries had hoped for. Meanwhile, LMICs and BRICS countries have grown more powerful, creating a multipolar international order.

As the US remains at the core of a technopolar world economy, its reduction of political power within the international order has led to declining interest in institutionalized track 1 diplomacy. Rather than building capacity to openly address double standards to reinforce and evolve international rules for long-term convergence, US foreign policy increasingly leverages American technological dominance to serve immediate national interests.

However, this retreat from institutionalized track 1 diplomacy carries significant costs. The resulting fragmentation of the international order tends to increase geopolitical tensions, fueling fear, doubt, and uncertainty among LMICs and allies alike. The US's withdrawal has left more room for the voices of authoritarian regimes. But due to a lack of capacity for international engagement even among the US' biggest rivals, the last decade of the international system has been marked primarily by a leadership vacuum. Greater US investment in diplomacy across all tracks is feasible with modest resourcing – without it, these channels will either cease to exist or calcify.

US diplomacy remains severely underpowered, mostly because of misconceptions around the current scale of diplomacy and the returns on additional investment. With less than 1% of the federal budget invested in foreign policy, the US only has around 8000 diplomats spread across 271 posts, and virtually none dedicated to frontier AI diplomacy. According to insider sources, decisions on US diplomatic engagement in AI governance are made by a handful of officials across the State Department, National Security Council, and Department of Commerce.

The intersection of frontier AI and geopolitics, especially focused on China, seems entirely neglected. Further, proxies whose voice is commonly understood to be independent of the US –

primarily LMICs – could help advance American interests cost-effectively. Unfortunately, their diplomatic capacity – especially on frontier AI – is even weaker.

American diplomacy – institutional and informal – has declined to levels well-below those of the 1950s (per capita). On top of maintaining existing channels, one way to advance US leadership in AI diplomacy is to open up new channels and pull the US government into them – a gap that non-governmental organizations can fill.

4. SI's contribution: opening and maintaining diplomatic channels on frontier AI to secure US diplomatic engagement.

SI's track record is an attest to the possible cost-effectiveness of diplomacy: with just a handful of dedicated people freed from institutional constraints, we have been able to counterfactually align key stakeholders who lacked either the mutual understanding, topical expertise, or resourcing to converge on the need for new international institutions: the International Scientific Panel on AI and the Global Dialogue on AI Governance.

Our current priority is to design and negotiate a strong UN International Scientific Panel on AI (track 1). The rationale for the panel is to foster convergence on global priorities by lending political legitimacy to the International Scientific Report on the Safety of Advanced AI. The panel's low financial cost (~20M$/year) paired with its value as a strong signal of support to non-developer states is likely to pay off in multiples. US support of the panel would provide fertile ground for further steps toward an international AI governance regime without handing control or boosting the influence of authoritarian states.

In addition to our work on institutional diplomacy, we're now exploring track 1.5 dialogues between US, China and proxy governments. These confidence-building measures could develop alternative pathways to the UN. We suspect most of the immediate impact would stem from serendipitously identifying the rare civil servants who can comprehend frontier AI dynamics and are sufficiently agentic to align parts of their government on the importance of international cooperation.

We are seeking feedback on whether and how we should leverage our independent positioning (we interface with North, South, West and East countries), neutral location (Switzerland) and policy track record (we've influenced international texts) to launch and fundraise for this work.

If you can spare an additional 5 minutes, please comment your takes on the following:

Where do you suspect SI could add most value in this domain? Some options, for example:

- Supporting the US government or EU in building a new kind of diplomatic corps, focused on frontier AI, leveraging our understanding of diplomatic channels and culture;
- Boosting our engagement with Chinese government representatives to gain more insight, leveraging SI's reputation as an honest, effective facilitator of AI diplomacy;
- Organizing meetings between US and Chinese nationals or useful proxies, leveraging Switzerland's neutral location and our connections to established mediators; or

- Coaching the senior diplomats and international civil servants we work with to drop out of foreign/international service and work directly on AI governance in existing or new orgs.