THE BROOKINGS INSTITUTION

WEBINAR

HOW WILL ARTIFICIAL INTELLIGENCE IMPACT SECURITY RELATIONS BETWEEN THE UNITED STATES AND CHINA? US AND CHINESE PERSPECTIVES

Friday, January 10, 2025

WELCOMING REMARKS:

COLIN KAHL
Sydney Stein, Jr. Scholar, Strobe Talbott Center for security, Strategy, and Technology, The Brookings Institution

ANDREW FOREST
Founder, Minderoo Foundation

PANEL DISCUSSION:

TING DONG
Fellow, Center for international Security and Strategy, Tsinghua University

LU CHUANYING
Nonresident Fellow, Center for International Security and Strategy, Tsinghua University
Senior Fellow and Deputy Director, Institute of Public Policy and Innovation, Shanghai Institutes for International Studies

CHRIS MESEROLE
Former Brookings Expert

QIAN XIAO
Deputy Director, Center for International Security and Strategy, Tsinghua University

JACQUELYN SCHNEIDER
Hargrove Hoover Fellow and Director, Hoover Wargaming and Crisis Simulation, Hoover Institution

MODERATOR: RYAN HASS
Chen-Fu and Cecelia Yen Koo Chair in Taiwan Studies, Senior Fellow and Director, John. L. Thornton China Center, The Brookings Institution

* * * * *

KAHL: All right. Good morning, everybody from the United States. And good evening to those of you who are joining us from Asia. For those of you who I don't know, I'm Colin Kahl. I'm the Sydney Stein Junior scholar at the Brookings Institution. I'm also a professor at Stanford University, where I run our program on geopolitics, technology and governance. And I'm a former undersecretary of defense for policy in the Biden administration. I'm pleased to welcome you to today's event, which explores how artificial intelligence AI will impact security relations between the United States and China. I'd like to begin by providing some background on the work that has brought us here today.

Since October 2019, the Brookings Institution's Foreign Policy program and Tsinghua University's Center for International Security and Strategy, CISSP have convened the US China track to dialogue on artificial intelligence and national security. I've had the privilege to participate in this dialogue as the head of the U.S. delegation for the past couple of years. Over those five plus years, we've held 11 rounds of dialogue. The next round will be held on the sidelines of the Munich Security Conference just next month. This ongoing series of meetings held in-person in third countries and virtually during the COVID pandemic has spanned two U.S. presidential administrations. It's brought together consistent teams of U.S. and Chinese experts on artificial intelligence and national security to examine where there is consensus and dissents on boundaries around uses of AI in national security.

Early on, dialogue participants identified the absence of a shared set of definitions for key AI terminology as a limiting factor for gaining precise understanding of each other's views on critical questions. The two teams work together to build a glossary of AI terminology that was published this past August and can be found on the Brookings and CISS websites. One of the areas of agreement that has come to light during our track two discussions is the aim to keep humans in control of nuclear launch decisions. And we're proud to say that this past November, at the meeting of Presidents Biden and Xi in Lima, Peru, Presidents Biden and Xi endorsed this principle. Our all-star lineup of panelists today will further explore the origins of this work, its key outcomes and the focus of these efforts going forward. You're welcome to submit questions for our panelists by emailing events@Brookings.edu or via X at Brookings FP using the hashtag #USChina.

This dialogue in this event would not be possible without the support of the Minderoo Foundation. I'd now like to introduce Andrew Forrest, founder of the of Foundation for some welcoming remarks. Andrew provides generous support to Brookings to make the work that we do possible. I'd like to reiterate that Brookings is commitment to independence and underscore the view that the views today expressed are solely those of the speakers. But with that caveat over you, Andrew.

FORREST: Thank you. Colin, Look, this all started in the middle, late last decade between a very senior leader in town and myself when I was breaking down the fact that I could be an untrammeled arms race to the death. That is, that it's way more dangerous than nuclear and that there is a keeping that power. We could actually use it for good, particularly in a military perspective, to engender and create peace on a long-term basis. You know, of course, everyone in this audience and Colin, your leadership's been critical here, all of you. Tsinghua University, Brooking, thanks for getting this message out. This is such important message that the broader thinking public know just how dangerous AI is as a tool in the military. But you correctly, it can actually be even more powerful than the nuclear piece we've enjoyed for some 80 years. That's relative peace, of course. I'm talking about no World War 1 or 2, but we haven't had that.

So as you all know, around 70 years ago, Dartmouth College machines were considered to be out of sync like humans in a way with AI. And of course, it's it is a very, very strong topic in China, you don't hear as much about it, but it's a huge topic over there. They're very advanced. It's obviously massive in North America. The data centers in impacting entire energy flows so big and it's made massive lifesaving contributions to brain cancer, which has always been almost untreatable, is now starting to be cracked thanks to air and in conflict zones like Gaza and Ukraine. The downside of AI and technology is on full display, where it strips away empathy, where decisions are made by machines on only information which it's been fed. And those decisions can be absolute rubbish. It's dangerous. It will never be as good as humans. And that's why I am. And I'm talking about humans for good, not humans for aggression.

So I am grateful, as pointed out, that I'm that we are pursuing global leaders. And we've had an excellent example in the United States President and the Chinese president reaching an agreement

that they're going to put in place. A safe system, a human system for the use of nuclear. And I want to say that the mantra behind all of this, Minderoo is funding our ability to make AI a friend of humanity, not a terrible enemy in the military is a simple four-word slogan, "no harm to citizens." It used to be military conflict. Terrible. You would wipe out citizens, you'd wipe out. And it's not even the decision makers we're facing being wiped out as well. Did we get paid to the Crimea war? U. World War One is some examples of this it brings those decisions right to the front. You've got now AI in the military ensuring that human beings, not artificial intelligence, retained.

As you've seen in this very far reaching environment between Biden and Xi, you've seen humans retain that decision-making authority. That was not the case in 2017/2018, when hair trigger advantages in timing would lead to a small advantage in military activity in warfare. And of course, one small area in AI and we're toast. So putting humans into that picture is critical for collaboration. Look, it's this progress we see in Biden, we've got human control over nuclear weapons. Decisions are on the right. Those two things are on the right track. But it came from the publication of a common glossary. Everyone thinks, well, what's so important about a glossary? Well, what is this? Is this a telephone or is it an instrument covered in the wrapping which has the ability to communicate averagely?

Those two lexicons in an instantaneous environment where decisions get made and they don't get understood because language isn't used to understand each other, isn't used to help us. That provides important context for actions. So getting that language right, getting that lexicon right, publishing that glossary we have important - I'm saying let's have the world follow this example is powerful nations can find compromise on humanitarian issues. But all I can say is then we should all be proud. And I'm proud of the teams, proud of the teams for making this happen. We've really been achieving something which many thought impossible and. And this is having a proper consensus on humanity's role in future conflicts. It's what I want to see. It's what we need to have the decision and the danger come straight through to the decision makers so that the decisions to send a country to war to destroy citizens.

I don't call the destruction of citizens collateral damage. I call that the destruction, the deliberate destruction of citizens to be murder. We've never had the ability to able to prevent that murder and warfare. I can give a set decision that brings that responsibility right to the front for action. So I just want to say thank you to you all. This is an incredibly important path we're on. Having the decision makers feel first that military targets decision making targets and not humans, not citizens of North America, of China, of the world are not collateral damage in warfare. That is the power of AI. That is the power which we want to release to make sure that in any military conflict, innocents, innocent citizens are completely protected.

It will also do what nuclear is done and help prevent an all-out war, because the decision for destruction comes straight to the person who makes the decision to destruct. So I advocate and congratulate all of you here to really push no harm to citizens. AI gives us that opportunity and possibility for the first time in history. So with that, let me just say thank you. Ryan Hass, senior fellow director of the John L. Thornton China Center at Brookings who will now moderate threat to today's program. I look forward to being with you. This is critical to humanity. It's critical, of course, to US military foundation. We need to see that decisions are made which protect citizens, which protect people. Don't just focus on the bad decisions of humans. Thank you very much.

HASS: Well, thank you, Andrew. And it's wonderful to have you spurring us forward, providing a needed sort of catalyst for accelerating our efforts. I also want to thank Colin for his leadership of our dialogue. It wouldn't happen without him. We're going to use the balance of our time today to pull back the curtain a bit on the track to dialogue that Colin and Andrew have been describing, to explore how it began, how it has worked, what we have learned, and what work lies ahead. We have an All-Star panel to help us think through and talk through these questions. I will briefly introduce them and then we will jump right into the conversation. So our first panelist is Dong Ting, CISS fellow. Dong Ting's research focuses on technology and nontraditional security. Thank you for joining us from Beijing. Shanghai. This fellow Lu Chuanying, who is research, focuses on cyberspace, governance and cybersecurity. Thank you as well for joining us from China.

We have former director and fellow of the Air and Emerging Technology Initiative at Brookings, Chris Meserole. Chris is currently the executive director of the Frontier Model Forum and is an expert on AI safety, international governance and global cooperation, among many other issues. Thanks for joining us, Chris. We have Hoover Institution fellow Jacqui Schneider joining us from the West Coast at 5:30 in the morning. Jacqui is the director of the Hoover Wargaming and Crisis Simulation Initiative and an affiliate with Stanford's Center for International Security and Cooperation.

Her research focuses on the intersection of technology, national security and political psychology, with a special interest in cybersecurity, autonomous technologies, war games and Northeast Asia. And finally, but certainly not least, we have Qian Xiao, deputy Director of CISS and also serves as Dean of International Affairs and vice Dean of the Institute for International Governance at Tsinghua University. Thank you for joining us from China this evening. So, Chris, let's kick off with you. When you were at Brookings, you were one of the key drivers behind establishing this dialogue over five years ago. Can you share with us why you helped launch us on this journey? What was the primary motivating factor in the goals behind the launch of this dialogue?

MESEROLE: Sure. And just first, let me start by saying thank you, Ryan, for putting this event together. And to all the participants that are on screen with us. It's a pleasure to see you all again. And it's been an honor not only to take part in this conversation, but I think now 11 conversations over the last 5 or 6 years as far as kind of where this track to dialogue started and where it came from. I think there's probably two core points that I would make and one is to pick up on a thread that Andrew for us alluded to earlier, which is this idea of an arms race and in other contexts with other technologies, the core idea of an arms race is something we've seen play out repeatedly in history where there's a new technology that has military applications. You know, states will invest heavily in the capabilities of that technology.

The other geopolitical rivals of states will kind of see, you know, some of their opponents invest heavily and then decide that they, too, need to invest heavily in the limit, what happens if you have you have kind of arms builds up built around arm buildups that can in in the limit lead and incentivize folks to actually engage in conflicts that neither party wanted in the in the first place. And so there's a

real danger that we've seen play out with arms races over time. And the challenge is when we started this dialogue in 2018 or so, the arms race kind of metaphor was really live. And it was partly to do with some of the breakthroughs that I'd come with AI, you know, at that point in time, we'd seen some really dramatic breakthroughs enabled by deep learning and narrow applications of AI, like the most notable of which might have been like AlphaGo, which was, you know, an AI system that enabled a model to outperform any human right, be the top human player and go. And it really caused a lot of consternation among not just the community, but globally.

I think everybody kind of assumed that that would be the kind of one of the hardest tasks for humans to do. And therefore the fact that I could solve it, it meant that the capabilities of AI were getting so good, so fast that we really needed to worry about some of the applications in high risk domains as well. And we've seen that capability trajectory only continue. And that kind of, you know, what we were concerned about in terms of growing AI capabilities has really played out over the last 5 or 6 years. What we were worried about was if you have kind of, you know, leading powers in the in terms of AI development, building up their AI capabilities without any kind of channel of communication that, you know, what was happening in terms of our development might lead to miscommunications and misunderstandings that, again, kind of in the limit, could lead to different kinds of conflict behavior that neither party really wanted in the first place.

And so, you know, at that time, you know, the last point I'll make about that, that era was it was a period in which there was, you know, very significant geopolitical tensions, very low trust between the U.S. and China in particular, and very limited communication in general, but especially on this issue. And so I think we thought, given the importance of the issue, given the importance of AI and given some of the rhetoric around an arms race that was happening in Beijing and in Washington, D.C., that it be worthwhile to have some channel of communication and to try and allow both sides to articulate a little bit more clearly, you know, what their intentions were, what you know, how they understood the technology and where they thought it might be going in the future. The second kind of point that I'll make is, you know, when we look back at other prior arms races, you know, AI is very dissimilar from the nuclear kind of space in many ways.

But one kind of core similarity is that, you know, when we looked at like how productive and constructive norms emerged in nuclear nonproliferation, it happened at the bilateral level first and then percolated out. And so we thought if we're trying to manage some of the worst-case dynamics when it comes to an arms race, which they are globally, it would be probably more productive to start at a bilateral level between the two leading superpowers that are investing in AI and see if we can make some progress there and then hopefully any kind of new norms or agreements that we were able to incubate at the bilateral level could then percolate out globally. And so, you know, thankfully, we've seen some evidence to suggest that that theory of the case was a credible one. Hopefully we'll see more in the future. But that's kind of the backstory of, you know, how we started on this journey. And it's been a real privilege and honor to take part in it in the years since.

KAHL: Well, thank you for saying that context, Chris. Xiao Qian, you also were present around the table for those early conversations in Beijing. From the Chinese perspective. What did you expect out of this dialogue? How is it performing relative to those initial hopes and expectations?

QIAN: Thank you. Good morning, Ryan. Thank you for having me here. It is I think it's really important to have both U.S. and Chinese perspectives on this project. And thank you very much for inviting the Chinese experts here. And you're right, that is that that I had a privilege of being in the dialogue at the very beginning and more share of the back stories here. I could remember very clearly that in July 2019, when she hosted our eighth War Peace forum and then our center, together with the support of our foundation, we had a special session, AI, Technology and AI governance. And I remember a lot of experts from around the world were invited, and they were the very first group of experts who studied AI, and they had governance and security and etc. So we had a very good discussion on risks and the challenges brought about by AI.

So after the session, I remember that you had the Brookings approach, the matter of when the Chinese former vice foreign minister, who was also the chairperson of the HAD, was at that time asking whether we can have a US-China dialogue on AI and national security. We both had a feeling that I could pose a huge threat to the global peace and stability in the future and agreed that the two leading countries, China and the U.S., should have dialogue and how we can deal with this risk. So

then the October that year we had a very good dialogue on AI and national security, which was the very first dialogue of our projects. And with the general support of me their foundation. I remember that at the very beginning of the dialogue, our main from the Chinese perspective, our expectation is to, as mentioned by Chris, to build a channel of communication and also to build mutual understanding and the trust between the two between U.S. and China, and avoid means of communication or misunderstanding between the two sides. So in order to achieve this, we had done three things, mainly in the past, in the five years.

First, to identify the potential safety and security risks regarding AI. Second, to establish degree of consensus and the rules regarding in military domain. And the experts explored issues such as target identification, compliance with international law, training data and system safety and etc. So the experts recognized each other's perspectives. New challenges posed by I in military domain. And also, we explored measures to reduce risks at different stages. And last but not the least, we tried to build a glossary of AI terminology so that we can gain the precise understanding of each other's views on critical questions, as, as mentioned at the very beginning.

So despite of all the difficulties the U.S. and the Chinese carried out of extraordinary work on establishing confidence building measures, experts acknowledge that factors such as declining mutual trust, the U.S. and between China and also the fierce technological competition between the two countries increased the uncertainty of establishing the confidence building measures between the two countries. Now, looking back, I personally believe, personally believe that we had achieved our original objectives and we have managed to build a mutual understanding and the trust among our experts. And we tried to build a dynamic community with AI in defense. Together, however, we have to admit that the geopolitical tension between China and the U.S. during these five years really had made our progress there are difficult and to a certain degree very and to some degree limited. So I hope maybe in the future we can have more, achieve more in our team. So I will stop here. Thank you.

KAHL: Well, thank you. Just to draw out two things that Chris and Xiao Qian pointed out, this dialogue started before talking about air was cool. You know, it's true that 2018, 2020. And it also

occurred during the first Trump administration. And so that's important. Points of context. Dong Ting, if I could turn to you next. You are a newer member of our dialogue with a very important expert contributor. What has been your reaction to the work undertaken thus far and in particular as an expert on nontraditional? Security. How do you think our work is or is not helping to bridge some of the security divides between the United States and China?

TING: Thank you, Ryan. Thank you for this question and my great pleasure to see a lot of you. Well, yes. When I joined the faculty of Divide two day to day one, I had already heard a lot about this dialogue project, but it was not until September 20th, 2023 that I became directly involved through the terminology Working Group, as Professor Xiao just mentioned, and I was really excited about this opportunity. So actually by that time I was no longer a news topic. As you guys do at the very beginning, five years ago. It had become a buzz word in international security discussions. So everyone was talking about risks. Everyone is talking about the need for governance. But when we look closer, that was a fundamental challenge, right?

So while everyone agreed on the need for risk management, but we kept running into a basic question what exactly were the risks we need to manage? So some of the risks being highlighted that time were essentially cyber age security concerns, just amplified by age capabilities. So what are the fundamentally new risks that I, as a transformative technology, brings to the table? So that is a question, right? So this is precisely why I found the creation of the terminology working Group so brilliant. So in that during the past five years, I'm so sorry to say that, but there was a time the substantive China-U.S. dialogue has become increasingly difficult. But our discussion about terminology actually create, well, what I would call conceptual meeting points, right? By focusing on how you decide, defines and understand key concepts.

We could begin to map out where our risk perceptions actually diverge. So it's not just about agreeing on destinations. It is about understanding why we define things differently. Just a very quick example. When We discuss terms like meaningful human control? We were not just trying to reach a consensus or definition. Instead, we are mapping out how our different technological traditions, security cultures and risk perceptions shape our understanding of AI. So this deeper understanding is

very important because the security challenges posed by are not just a technical problem. They're deeply intertwined with how we think about technology, throw in society and the security. So it was particularly but even more about this approach in that it allowed us to move beyond the binary thinking of cooperation versus competition. So our terminology group, AI, we were discovering that even within our definitions for understanding why we do for just a can create some new policies forward for further dialogue.

So sometimes this difference is actually a review. We have showed shared concerns, which is to view this through a different culture and strategic perspective. So I believe this kind of foundational work, meaning understanding each other's conceptual framework, is very important for building more resilient security dialogues that come by the broader political tensions. So in this sense, I think our dialogue is doing more than just a bridging current security devices, is helping to create new ways of thinking about how we two major powers can engage on complex technological challenges, while differences in perspectives on natural right. This kind of system dialogue shows us through careful communication and mutual understanding, we can gradually reduce misperceptions and expand new spends for positive interactions. Let me stop you.

HASS: Thank you, Dong Ting. I think it was really important that you highlighted the fact that this dialogue is not an exercise in finding the lowest common denominator. It's an exercise in trying to really clarify our understanding of each side's perspectives. So thank you for really trying not to the surface. Jacqui, you have been our resident expert on Wargaming. Throughout this dialogue. You've crafted many scenarios that our dialogue teams have helped use to clarify understanding of concepts and terms, help our audience understand how war gaming has been used to sort of spur the discussion and some of the learnings that these efforts have uncovered.

SCHNEIDER: Yeah. So I think we, we started with games really in this very one-dimensional scenarios. So the first attempt to try and introduce a gaming element to the dialogue was introduction of really relatively simple hypotheticals. And the idea was that these even these simple hypotheticals and. Serious could bring us out of our routine and out of our scripts and out of kind of rote discussions into more of an immersive conversation. And I think in general, the idea behind scenarios and games

as mechanisms in order to do diplomacy and to better understand the dangers the future is, the games allow us to move out of the reality that we live in and into the most dangerous forms of scenarios. And from for us, from the accidental use of AI enabled conventional weaponry to AI enabled disinformation and eventually actually to discussing campaigns and the intersection between AI and nuclear use.

So, you know, by presenting a game and a scenario and a hypothetical, it allows us kind of the space to be able to discuss something that would be otherwise relatively difficult in a really and a, you know, kind of fixed and more conventional conversation. I mean, I think the other things for us, we started with these really simple but sometimes very dangerous scenarios. I think in our first attempt use scenarios, we included that AI and nuclear use. Like, boy, are we ready to talk about a nuclear. These are kind of the biggest and scariest things. And we found actually we are by creating scenarios and hypotheticals, we were able to discuss concepts that might otherwise have been really concerning and difficult for us.

So over the years, those scenarios, we went back and forth with, you know, sometimes the American side presented the scenarios, sometimes the Chinese side presented the scenarios to moving to actual game situation where when we're in person, we're able to combine teams and play on the same team in order to understand the dangers of an inadvertent escalation from an AI enabled and mistake. And in a conventional wartime scenario, the thing is that these games allow us to create scenarios that places out of us versus them. And always when I designed the scenarios, I was very careful to say I couldn't tell in the scenario, is this China or is this us? Instead, it was, Hey, I'm going to create a scenario where an inadvertent thing or an accident happens and it really could have been either side and allow us to think through like neither side wants this inadvertent or accidental thing to happen.

So how can we think through those dangers? So I think and what the games provide is an ability to think about the dangers of the dangers that the countries share. So you can decide games that are primarily about competition, but instead we're using games within this dialogue as a way to show where there are potential kind of similarities about how we understand the dangers of the technology,

of the technology. And I think the idea is not to prognosticate the future, but instead to use games as a way to understand the most dangerous futures and then to find recommendations about how to avoid these most dangerous outcomes. And I think I've also focused here a little bit. This is a pretty optimistic view of the role in which games can create trust and allow us to talk about dangerous situations that might otherwise be difficult within a dialog.

But I think the other thing the Games reveal and have revealed in the process of this dialogue is, is where we actually don't agree or where we misperceive each other. And I think end the identifying potential friction points. And, and then finally, I want to say, you know, we kind of have gone back and forth with these scenarios. Sometimes the American side determines what scenarios we're talking about. Sometimes it's the Chinese side. And actually, that process of creating the scenarios it both reveals to the designers, hey, what do we care about and how does like developing the scenario in the game help our individual side understand where we think there are dangers that we'd like to discuss, but it also helps us perceive kind of what the other side views as the most dangerous.

And I think sometimes even the asymmetry between kind of what we proposed as different scenarios really reveals a lot about what each about asymmetry is and what where each side saw the biggest dangers. So I'll go ahead and stop there, but happy to answer more in question and answer.

HASS: Great. Wonderful. Lu Chuanying, you've been a participant in these war games that Jackie was describing. Obviously, United States and China have different political cultures, different ways of working through some of these challenges. Can you help us understand from a Chinese perspective, how have the games worked for you and for the Chinese team? But also a second question. As a cybersecurity expert, are there any useful lessons to be learned both in the context of the dialogue and our mutual learning around? These issues that could be applied to cyberspace already.

CHUANYING: Yeah. Thank you, Ryan. And also talk about the scenarios I showed. Thanks to Jackie and also the Chinese, the pundits and senior counsel with the help to give us a lot of very good case. And I think during that, I think the most impressive part is the discussing, right.

We have a very broad, diversified background, people sitting in the room and discuss the different play, the different rules and discuss the different situations that really help me to better understand how the conflicts between country might play out in the future scenarios with artificial intelligence. So that that's also helped me to better understand it, that the advantage and disadvantage of using AI in the battlefield, right. At one hand, we have brought us more high-quality information very fast. But on the other hand, it also will create some kind of uncertainties which will lead to kind of the escalations. So that is dangerous. Very dangerous.

So there are also some lessons I learned from discussing this. First, I remember during the discussing, nobody want the things get out of the control because we don't know how the thinking goes with the artificial intelligence. Right. So, almost all the people empathized a lot of times of the humor in the loop. We need to keep the AI under the control of human beings. And the thing now understanding is that we still have a long way to go to understanding using AI in the battlefield because there are a lot of the uncertainties. So it is crucial for us to refrain ourselves in developing military AI and also using them in the battlefield. And the second point is, I think it is an urgency for us to for the global community to work hard to establish some kind of the rules and norms in the field because of the know-how for consequences of using AI.

So that's my response to your first question. And second question really are a good question. Help me to think you make a comparison between the Cyberspace governance and AI governance. I think that for their cyberspace governance, we spent a lot of time on the A state behavior website, a lot of the responsible state we have here in cyberspace. But the problem is, without the technology, it's hard to verify they have you. So people are turn to accuse each other. You're not, you know, follow the norms. You are not responsible state. And therefore, the governance people are more concerned nowadays about the safety, which I think is a very good start because it's more technical and also, it's more easy to find the common ground because none of us want the AI out of control. We all want to have a safe AI system. So that I think that the lessons that the cyberspace government should learn from the air to pay more attention to the technical part. Yeah, I'm going to stop here, right.

HASS: Great. Well, Chris and Xiao Qian, one of the themes that has sort of emerged over the course of our conversation is that when this dialogue started, it was a very sensitive topic and a period of stress in the US-China relationship. And it was not clear at that time whether or not the two sides would be able to productively advance discussion and really sort of clarify sensitive issues related to the use of AI and national security. Here we are five years later. I want to give you an opportunity to reflect both on anything that you've heard from Jack here to training and don't take, but also on the journey thus far. Are there any useful lessons that could be learned both in the context of the dialogue and in the mutual learning on air and national security? Chris, I'll start with you and then turn to our discussion.

MESEROLE: Good. Thanks, Ryan. I would say that there's been a lot of lessons learned. I think especially when you're thinking about how to build a context of high trust within, you know, low trust environments. There's some, I think, tricks of the trade that we learn from and developed through trial and error within this context that are probably useful to discuss. And I would put them in kind of three buckets. The first is just the ground rules that we operate under. I think we're obviously having a very public conversation right now. But when we started, we kind of very explicitly said we weren't going to be that public, at least for the start. And I think that, you know, a few years, even before we actually published anything publicly about the dialogue and I think allowing kind of taking the pressure off, you know, the conversations in the sense that I think we all understood that this was like an informal or relatively informal kind of private setting in which to discuss these things really helped to build more trust.

The second was just some it sounds very basic, but just providing side like especially for the in-person meetings, like having side kind of availability of side conversations and dinners to just build, you know, personal human relationships. It's an it's a really tricky thing to say, but it's incredibly powerful to have like those moments to build up high trust. And I think especially given what happened with Covid and the way that we had to shift to virtual, I think it was really very much to our advantage that we had already a baseline of kind of in-person relationships that had been developed.

And then the third kind of bucket of things that I think we did as we found kind of intellectual mechanisms to identify, to hone in on the actual issues divorced from some of the broader geopolitical dynamics that I think, you know, Jackie and others have done a great job laying out kind of, you know, why it's so valuable to have vignettes or war games or kind of why it's, you know, you talk thing was mentioned in kind of the glossary like why it's so important to have like those kinds of mechanisms to focus on. The one additional thing I'll add in as one of the things we did, especially early on, was ask each side to put in writing some of their own understandings of different topics. And I think it was, you know, so each side would kind of circulate memos in advance of the initial rounds of the dialog. And the ones that I think in my view were most impactful in some ways were actually putting in writing our understandings of the failure modes of AI.

Because I think when we talk about the AI arms race, there's a unique wrinkle to AI, which is if both sides are convinced that the other side is unaware of some of the failure modes and is not properly attending to putting in place the safety appropriate safeguards, there's an incentive to also cut corners on safety and security of these systems as well and just kind of rush their development. And I think one of the big takeaways of how we structure the dialog, both enabling kind of these conversations to happen in a high trust way, but also just learning from each other was those early memos we wrote where we both kind of put in writing our understanding of kind of, you know, how I operates. And it took out the logic of, well, you know, these people, you know, these are our counterparts don't understand the technology.

And so therefore, we don't need to put in place the safeguards either. And like that, I think we were able to successfully counter argue some of the logic and arguments that were happening in that vein for proponents who wanted to just kind of put the gas, to put the pedal to the metal, so to speak, in terms of rolling out AI military systems before they were really fully baked. So and I'll stop there before I go down too much of a tangent on that. But all told, I think all of those mechanisms really helped to bring out like a really high trust environment in which we had these conversations.

HASS: Thank you. Xiao Qian?

QIAN: Hey, thank you, Chris. I think you have covered most of the points that want to mention. Actually, I think I should first analysis analyzed the three factors that why we could continue the discussion amid the political difficulties. I think the first is that the both teams believe that this topic is extremely important with the fast development of AI technology, the society is getting very worried about the risks and the challenges, especially in the military domain. So that's why the teams and every expert is very consistent in continue our joint research, even though we have a lot of difficulties ahead. The second thing is that we both believe that is very important for both China and the U.S. to have a dialogue on this important issue.

The two countries, I mean, the most important bilateral relations that is so important that for the world and that we cannot afford any miscommunication or mis understanding between the two countries on this issue. Although the government may not be ready to discuss goes about this very directly right now, the academy people or the think tank people should go ahead and get ready for the future. Risks and challenges, so we should do some research beforehand. The third factor is that we have precedent. That is, we know that the P5 countries, they have published their joint glossary on key nuclear atoms in 2022 after ten years of difficult discussions. So they had a very good example, and they could reach consensus on such sensitive issues. Why not challenge the U.S., reach some consensus on the joint glossary? So we will quite be encouraged by their efforts. And then the last two points about the lessons we have learned from our dialog.

The first one is that we should have an academy community for this because is a totally new area. So we have to build a community for this. Here in China, we have a pool of experts of more than nearly 30 and more than 20 there. They may not be able to attend our discussion every time, but they made contributions to our project one way or another. And we had internal group meetings from time to time and we learn new things and we try to discuss with each other about the new developments of our technology in the military domain. So which is really important for our research. And the second thing is that the alignment of objectives of the dialogue is very important between China and the U.S. teams. Our dialogues took place amid the very escalating strategic difficulties and the decline in mutual trust.

But our scholars generally believe that the governance of air enabled weapons, particularly autonomous weapons, was very important and was paramount. So both teams believed that it is very important to discuss about this, no matter how difficult the bilateral relationship is. We should keep the channel of communication open in this area to discuss about those very difficult issues. So we admit that the big trend of bilateral relations have impact on our dialog. Sometimes the more sanctions coming out, then it is more difficult to for us to have dialogs. But we still keep working hard and we also hope that in the future we can and have a more smooth development of our dialogs. I will stop here. Thank you.

HASS: Thank you, Xiao Qian. I'm going to commit as quickly to another topic. Artificial general intelligence. My impression has been that our Chinese colleagues have been a bit more skeptical about the potential development timelines of AGI or even whether AGI will be reached at all. Jacqui, can you help us sort of situate this this discussion a little bit? What is artificial general intelligence? Why would it be so significant if it's realized?

SCHNEIDER: Yeah. And I think, you know, here we there really is kind of an alphabet soup about kind of the words that we use to talk about the growth and development of artificial intelligence, which I mean, which is why it highlights why creating that glossary is so important. And so, I mean, quite simply, the idea behind artificial general intelligence is that this is a progression of AI in which air is able to either equal or actually surpass or better human cognitive capabilities. So in instead of it being narrow AI, where we're using AI for specific tasks, think, for example, and the use of AI to help guide a missile seeker head. Right?

It's relatively narrow task. Instead using an AI for larger kind of decision making, as you know, potentially a replacement for decision making on very large strategic questions. So in the context of these discussions and you know, what we think about as existential risk from AI, the real concern is, is AGI going to replace human decision making for the most important decisions that occur within the strategic and interactions? So is it going to replace that human decision making about the beginning of a war and about the escalation in violence and the escalation of violence potentially all the way up to nuclear war? And I think the concern is, if we are at that point, then are the machines making the

decisions about when and why to states like the US and China go to war and the extraordinary amount of violence that could ensue because of that.

HASS: Thank you. And Dong Ting, and/or, Lu Chuanying, from your perspective, why do you think there is a gap in expectations between the U.S. and Chinese teams on this question that Jacqui is describing? And what do you think that tells us?

TING: Well, thank you for this interesting observation, actually, based on our technology discussions. I've noticed this. There is a gap, right? But just to my own understanding, it just stems from different philosophical and practical perspective on the intelligence itself. So from a philosophical perspective, I believe this difference reflect the distinctive epistemological traditions in both countries. So the Chinese approach to intelligence is deeply rooted in our empiricism, right? Not in a limiting thing, but in a fundamental basis of human knowledge. So there is a Chinese saying you tend to you "do learning." Wisdom, by meaning you learn your wisdom from setbacks. So which perfectly captures our view of intelligence as a process of learning through experience, by learning, by doing and consistently approaching, but never fully reaching absolute truths.

So this contrasted with a Western philosophical tradition that distinguishes between physical knowledge gains through experience and metaphysical knowledge that transcends experience. So this distinction, I believe, partly explains the different attitudes towards AGI. So the Western approach, with its emphasis on absolute translated knowledge, might naturally lead to a more confidence achieving AGI as defined a more achievable goal. Well, in contrast to the Chinese perspective, turns to see intelligence as an endless process of accumulation and refinement. So making us put more cautious about claims of achieving a complete AGI. Then from a practical perspective. I just want to emphasize that AGI as a technical goal naturally involves multiple high potential technical possibilities. This is true for any country pursuing AI development.

What differs, I believe, is how we frame our immediate objectives and developmental approach. Many Chinese researchers and companies here take what I would characterize as a more application focused approach rather than targeting AI or debating AI's potential to surpass human intelligence.

They just are focus on developing and improving a practical AI system. So this has led some observers to describe China's approach as pursuing AI generated visual intelligence rather than AGI. That being said, this pragmatic focus doesn't indicate skepticism about eyes potential here in China. In fact, many Chinese professors and experts, they believe we might see breakthrough development over the next five years or so. So the dif, the key difference, losing their views on AI's is how we will get there. I think many are not convinced by the current dominant approach, particularly the emphasis on large language models is necessarily the only path towards AGI. So and there is also a capital down here to consider. Given the maturity and the structure of the venture capital markets in both of our countries. Here in China, many startups. Entrepreneurs, the need to focus more on integrating AI into existing business models and so that they can deliver memorable for value, meaning making money very quickly. Right? Well, our U.S. colleagues might have more latitude to pursue longer term, spectacular research thanks to a more established venture ecosystem. So but I just want to say that these different perspectives are not obstacles. They are not bad. They are valuable contributions to our global discourse on AI development. Let me stop you right there.

HASS: Chaunying, do you have anything that you want to add to that?

CHUANYING: And I want to go back to Jacki and explore this project. There's just one point. I'm not representing the Chinese or I just my personal understanding. So I believe the current technology technological path will not lead to the artificial intelligence AGI, because what I have been thinking or what I have been using is basically the AGC. Right? It's very good to generate text image, but not the way you need to. That basically is still the now the. It's still based on the scouting lore and the big data. It could know in the in the future it will now increase its performance, but not come to a way to solve it. Just like people saying it, People turn to overestimate the short-term breakthrough of the technology, but underestimate the long-term breakthrough. But however, we still know how long-term technology breakthrough come and whether it will bring to us to the AGI. So that should have.

HASS: Thank you, Jackie.

SCHNEIDER: Yeah, I just want to highlight that, you know, AGI, whether something is AGI or not, AGI is really not a binary right? It exists on a spectrum. And what you're hearing here is a discussion or about where we currently are on that spectrum and how far we can get on that spectrum. But in some ways, that's kind of an academic conversation. I think what's important to emphasize is that it's the perception that states have of where a guy is on the spectrum. That is important because it is that perception that defines how states are willing to trust AI for large strategic decisions. And so I think in the future, it's helpful to get beyond like is this guy is not a guy. And instead and at what point do you think the decision makers think it's AGI, and therefore are willing to delegate decision making to AGI?

HASS: Interesting. I have a few questions that we've received from the audience. I'm going to put one on the table, but I'll also give you a chance to weigh in. If anyone shouts or Chris, you want to get in on the conversation as well. The first question is from Sydney Friedberg, who is a contributing editor at Breakingdefense.com. He asked, how are the U.S. and China taking similar approaches to military use of AI? In what ways might there be diverging? So this is an open question. Well, welcome. Anyone to jump in on it?

SCHNEIDER: Crickets. I, look, I don't think anybody on this panel or in this dialogue can speak directly for either side's military. And I say just from the US perspective, I think the US has been and pretty for forward leaning on thinking about the policies, about the ethical and appropriate use of AI in conflict and not just in the adoption and use of the technology, but also in the development of the technology. And so I think and throughout the dialogue there has been a bit of an asymmetry between the US has already created policies that it has distributed and publicly discussed. And whereas we're still waiting for those kind of public commentaries to come out from the Chinese side.

HASS: Right. Next question we received from Ron Axler, who's a senior research director at the Hackett Group. And Chris, this may be a good one for you to start off on and others can jump in as well, since it touches on something that you raised. Are we destined to be in in an ongoing arms race between the United States and China for air superiority?

MESEROLE: It's hard to I mean, it this is one of those kind of hard answers where it really depends on the definition of an arms race. I think there will be I think there'll be continued investment in capabilities on both sides. I think the more important thing is what are the perceptions that each side has of what the other is doing and do we feel like we understand what the intention and goal of that activity is? And if we have channels of communication like this to help manage that information flow there, I think we'll end up in a better place than if like the real worst-case scenario is we have a really disconnect. There's little communication and both sides are going all in on it in a way that is concerning to the other. And there's just not much communication about, you know, what the angles are or, you know, there's not much understanding about how the technology might be used within each side's military. So I'll pause. We could have a whole separate, I think, like hour long discussion about that. So I'll just I'll pause here for now.

HASS: Would anyone on the Chinese side like to weigh in on this question?

CHUANYING: Yes. So I think I agree with what Chris said it well. There's strong connections between the, you know, artificial intelligence and the military. Right. The all the government to have the no interest or to do this to develop the artificial intelligence in the military domains. But could that be defined as the arms race? It's still not clear, at least so far from China's side, are very concerned about the impact of the global or China/U.S. arms race. So it's very self-refrained from that trend. Yeah, but.

QIAN: I also want to add one thing here, I think. As for the arms race scenario, I think if the countries focus more on competition instead of cooperation, then it is very easy to go into the arms race scenario. And whenever you use you have development in technology, then it is used in the military. Then it is very easy to be into that situation. And then about the age, I think I also want to mention that even within China, we have a lot of debates on whether we will have an AGI in the near future, because some scientists I strongly believe that we are going into that direction. But there are experts like training. They don't believe that we will have that in the very near future. So it is always a very fierce debate here in China as well.

HASS: OK, well, we have one-minute left, so if anyone has a parting shot, anything that they want to get on the table before we close. Feel free to speak now or hold your peace for the time being.

SCHNEIDER: I just want to say, just echoing some of Andrew's early comments and hooking on to the arms race. Not all air development is an arms race and not all air development is dangerous. AI for medical and human good is not a competition that is good for the world, right? What we know from political science is that arms races are when we're using technology to create uncertainty about states intentions or that increases the propensity for offensive campaigns. Those are two characteristics that we can discuss within this dialog, But I think it's helpful as we as we leave this conversation that not all is bad and that the point of these conversations is to find ways in which air can continue to develop to help humanity and for the common good without creating these inadvertent and uncertain characteristics that create the chance for war.

HASS: Right. Well, I think that's a perfect note to end on. And I'm thankful to all of you for sharing your insights, your expertise, and also your perspectives on the arc of this dialog. I think that the past hour has demonstrated that it is possible for U.S. and Chinese experts to talk about very difficult things. You know, we broach the topic of war between our two countries and how I could be used in it, but in a way that allows us to hear each other. So I thank you for sharing and sharing your insights. This is an ongoing project that will continue to move forward in efforts to find areas of consensus and to census around acceptable uses of artificial intelligence and national security. We look forward to coming back to our audience in the future with additional updates as progress merits. But for the time being, thank you to our panelists and goodbye.