

Forecasting the AI and Nuclear Landscape

Alexa Wehsener, IST

Leah Walker, IST

Ryan Beck, Metaculus

Lawrence Phillips, Metaculus

Alex Leader, Metaculus

September 2022

The Institute for Security and Technology

The Institute for Security and Technology (IST)'s mission is to bridge gaps between technology and policy leaders to help solve emerging security problems together. Our non-traditional approach has a bias towards action, as we build trust across domains, provide unprecedented access, and deliver and implement solutions. With a network of experts in Silicon Valley and influential policy makers in national capitals, we are the preeminent US West Coast institute positioned to outpace security risks and facilitate solutions before they become catastrophic to society.

Our areas of focus cover a broad portfolio: from cybersecurity and information warfare, to the rise of digital authoritarianism, to the role that machine learning (ML), artificial intelligence (AI), and other emerging technologies play in shaping national security policy and political environments. Our work is based on an applied research approach, seeking practical solutions to well-defined problems and iteratively testing and evaluating proposals before launch.

Metaculus

Metaculus is a forecasting technology platform and aggregation engine that convenes an active global forecasting community, hosts forecasting challenges and competitions, and contributes to innovation in the field of forecasting science.

Scoring forecasts according to their accuracy allows Metaculus to build statistical track records for each forecaster and for the community as a whole. The most accurate and well-calibrated forecasters are given more weight in the optimal aggregation of the Metaculus community's predictions, known as the Metaculus Prediction.

For projects like this one, Metaculus engages a team of Pro Forecasters, who come from a variety of backgrounds, but are connected by the common characteristic of having excellent track records. Indeed, Pro Forecasters have proven their exceptional forecasting ability—they represent the top 2% of forecasters on the platform out of thousands of competitors worldwide. Metaculus projects leverage the talents of the Pro Forecaster team at times when accuracy and calibration are of the utmost importance.

Table of Contents

Table of Contents	3
Introduction to the Challenge	4
IST-Metaculus Partnership and Methods	4
Landscape Forecasts	5
AI-Nuclear Integration	6
Measures of U.S.-China Tensions	8
Nuclear Use	9
What's Next?	10

Introduction to the Challenge

New technologies often have the effect of increasing tension in international relations. The interaction of novel machine learning (ML) capabilities with nuclear weapons systems, and the effect these interactions might have on individual nuclear weapon state actors and on the relations between them, is particularly concerning. As novel technologies mature and ML and artificial intelligence (AI) techniques are integrated into complex military command and control systems, diplomatic confidence building measures (CBMs) can potentially mitigate some of the risks associated with these novel advancements.¹

To ensure that diplomatic efforts are informed by technical expertise, transparency and trust-building efforts must revolve around collaboration between senior decision makers and the AI safety and alignment research communities—especially regarding the command, control, and communications systems of nuclear weapons (NC3). With rapid advances in AI and ML, the opportunities for integration are growing, and barriers to integration are significantly lowering. Understanding when and how nuclear weapon states may seek to integrate AI/ML into their NC3 systems is the first step toward preparing for changes in strategic stability spurred by technology adoption.

IST-Metaculus Partnership and Methods

The Institute for Security and Technology (IST) and Metaculus have launched an initiative that engages Metaculus Pro Forecasters to gauge the probability of success of potential CBMs applicable to the world's riskiest nuclear relationships. In the initial phase of this partnership, we are focusing on useful intervention points in the United States-China relationship. The United States and China are in a strategic competition for advantage in multiple techno-industrial sectors, including cybersecurity, AI and ML, quantum computing, and renewable energy. Forecasting provides the opportunity to identify potential areas of risk and opportunity in this strategic competition, and to understand the trends that may shape both desired and unwanted futures, helping decision-makers navigate the complex environment at hand. As a discipline, forecasting also imposes precision requirements that can surface policymakers' assumptions and sharpen their vision of plausible futures.

To ground the forecasting process, IST first imagined worst-case scenarios and extrapolated the trends that would have to converge to make such scenarios happen. IST and Metaculus then worked together to develop a series of questions for forecasters to explore the probabilities and drivers behind each trend. This project draws on nuclear and policy subject matter experts to inform Pro Forecasters' predictions. Framing questions for our work included: When and where

¹ Diplomatic confidence building measures (CBMs) act to address, prevent, or resolve uncertainties among states. CBMs are designed to prevent wanted, and especially unwanted, escalations of hostilities as well as build mutual trust on the international stage.

could AI/ML be introduced into NC3? And at which points could such technologies make a clear difference in conflict escalation or management? These predictions will provide a broader understanding of outcome probabilities in the context of US-China geopolitical competition and nuclear weapons modernization.

The remainder of this report explores the forecasts produced by the Metaculus Pro Forecasting team on questions within the following categories:

- *AI-Nuclear Integration*
- *Measures of US-China Tensions*
- *Nuclear Use*

Landscape Forecasts

In the initial stage of this collaboration, the primary objective is to assess the risks of escalation between the U.S. and China, including by the integration of AI into NC3. We developed questions across three categories, listed below, focusing primarily on a five year forecasting horizon. These questions, along with the probabilities estimated by the Pro Forecasters, provide a valuable starting point for quantifying the risks and opportunities in the AI and nuclear landscape.

AI-NUCLEAR INTEGRATION

1. Will the U.S. and/or China have integrated an artificial neural network into their nuclear command, control and communications systems before 2028, according to publicly available information as of 2038?
2. When will it be reported that the U.S. and/or China have fielded weaponry on an unmanned naval vehicle?
3. What will the share of high impact AI publications from China be in 2027?

MEASURES OF US-CHINA TENSIONS

4. Will China attempt to undertake kinetic actions in contested territory before 2028?
5. Will there be a military casualty caused by an air or maritime incident between U.S. and Chinese military forces before 2028?

NUCLEAR USE

6. Will China formally abandon their “no first use” nuclear weapons policy before 2028?
7. Will China conduct a test detonation of a nuclear weapon before 2028?

Below we discuss the median predictions of the Metaculus Pro Forecasters.² Some questions are binary, meaning they have a “yes or no” outcome, while others are numeric or date ranges, producing full probability distributions across possible outcomes.

In addition to the forecasts, we provide a brief overview of the reasoning forecasters used to arrive at their predictions. The team may have a wide range of reasoning, but we highlight a few common themes or interesting arguments.

AI-Nuclear Integration

Forecasts in the *AI-Nuclear Integration* category explore the likelihood of integrating artificial neural networks (ANN) into nuclear command and control (NC3) systems. Additionally, we believe that the use of remotely or autonomously operated naval vehicles armed with weapons by either the U.S. or China may signal greater openness toward AI integration and could lead to greater tensions between the two countries, so this question is also examined. Finally, the share of peer reviewed AI publications coming from China will be an important measure to assess relative investments in AI, especially considering the close connection between some Chinese universities and the Chinese military.

In assessing ANN integration into NC3, forecasters were asked whether the U.S. and/or China will have integrated an ANN into its NC3 systems by January 1, 2028, according to credible public reporting as of January 1, 2038.

Specifically, the question asks whether an ANN will be integrated as an essential component (including in conjunction with, or to be verified by, legacy systems) in missile trajectory planning, trajectory optimization, target selection, target prioritization, target optimization, de-noising, modulation, or anti-jamming, but excluding AI used in wargaming simulations.

Forecasters anticipate that neural networks may be integrated into NC3 in limited ways in this time period. If so, de-noising and trajectory planning may be potential implementations where neural networks could be integrated alongside parallel, non-AI systems that could provide additional verification. Forecasters expect that China is more likely to integrate neural networks into NC3 systems than the U.S., mainly due to perceptions that China (1) has made AI a strategic priority and (2) has fewer established legacy systems making adoption of new technology easier. Integration by the U.S. or China may be difficult to confirm as public information would more likely be available for larger integrations as opposed to integrations at the system subcomponent level.

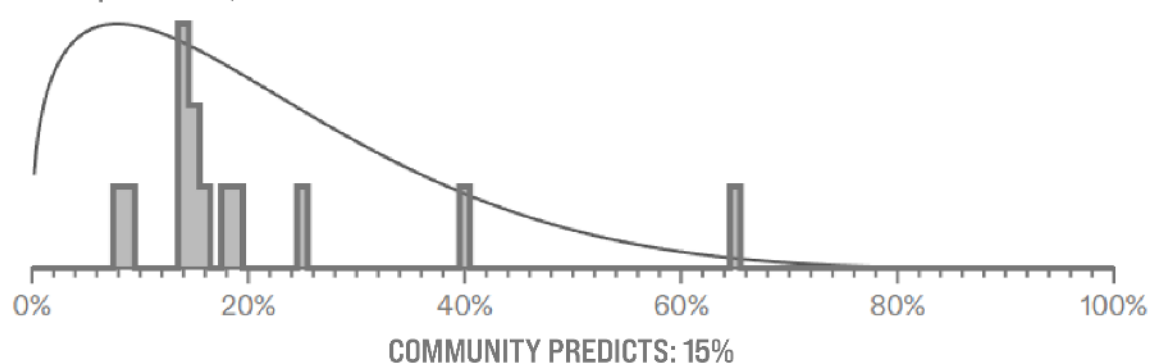
² Metaculus reports the median prediction from the Pro Forecasters, or in other words the middle value of the Pros’ probabilities and probability distributions for each scenario. For general, public questions and tournaments hosted on Metaculus, the Metaculus Prediction is typically recommended, since it is weighted by the most accurate forecasting track records and optimized with machine learning. However, with the team of Pro Forecasters, Metaculus has strong evidence that the individuals are well-calibrated, which makes the median forecast a reliable choice.

Overall the Pro Forecasters estimate a 20% chance the U.S. implements ANN into NC3 by 2028 and that we have public information about it by 2038 and a 15% chance of the same for China.

For example, one forecaster estimates a 60% chance of US implementation with a 35% chance it would be publicly revealed, compared to 70% and 15% for China, respectively.

Will China have integrated an artificial neural network into its nuclear command, control, and communications systems before 2028, according to publicly available information as of 2038?

As of September 19, 2022



To resolve these questions, Metaculus will use credible, publicly available sources and reporting. For example, while some forecasters think there is a chance that each country has already fielded an unmanned naval vehicle armed with offensive weaponry in a non-testing, non-exercise environment, there has not yet been public reporting to confirm it.

As evidence that technology is unlikely to be the limiting factor in fielding weapons on unmanned naval vehicles, forecasters point to Russia's allegedly autonomous nuclear torpedo, Poseidon,³ which is reportedly in development, and Israel's Protector USV,⁴ which has been in service since 2005 and can be fitted with a Mini Typhoon⁵ remote-controlled weapon station. In general, forecasters expect China to be slightly behind the U.S. in regards to such vehicles as, historically, the Chinese military is behind the U.S. for tech development and integration.

³ H I Sutton, "Russia's New 'Poseidon' Super-Weapon: What You Need To Know", Naval News, March 3, 2022, <https://www.navalnews.com/naval-news/2022/03/russias-new-poseidon-super-weapon-what-you-need-to-know/>.

⁴ "Protector Unmanned Surface Vehicle", BAE Systems, https://web.archive.org/web/20070510153843/http://www.baesystems.com/ProductsServices/protector_unmanned_surface_veh.html.

⁵ Yaakov Lappin, "Rafael launches improved Typhoon weapon station", JANES, November 2, 2020, <https://www.janes.com/defence-news/news-detail/rafael-launches-improved-typhoon-weapon-station>.

Predictions from Metaculus Pros indicate a median forecast of September of 2025 for the United States, with a 25% chance the U.S. fields weaponry on unmanned naval vehicles by May of 2024, and a 75% chance it happens before November of 2027. For China, the Pros arrived at a median forecast of May of 2026, with a 25% chance it happens before October of 2024. The upper end of the question range is January of 2028, and forecasters estimate a 29% chance it happens after that date for China.

In predicting the share of peer reviewed AI publications coming from China, **forecasters expect China's share of high impact publications on AI to grow to 30% in 2027, up from about 24% in 2021, due to China's significant focus on AI.** Several forecasters suggest that there's likely a ceiling on growth in the share of high impact⁶ publications coming from China, due to China's demographic change (i.e., fewer young researchers) and increasing competition from other countries (notably the U.S., European Union, and India) as spending and effort in AI grows.

The question about AI publications uses Scopus data provided by OECD.AI,⁷ specifically the percentage of high impact peer reviewed publications from China relative to the rest of the world. Scopus defines "high impact" as having a Field-Weighted Citation Impact (FCWI) of greater than 1.5, meaning that high impact papers have 1.5 times more citations than average in a given field.

Measures of U.S.-China Tensions

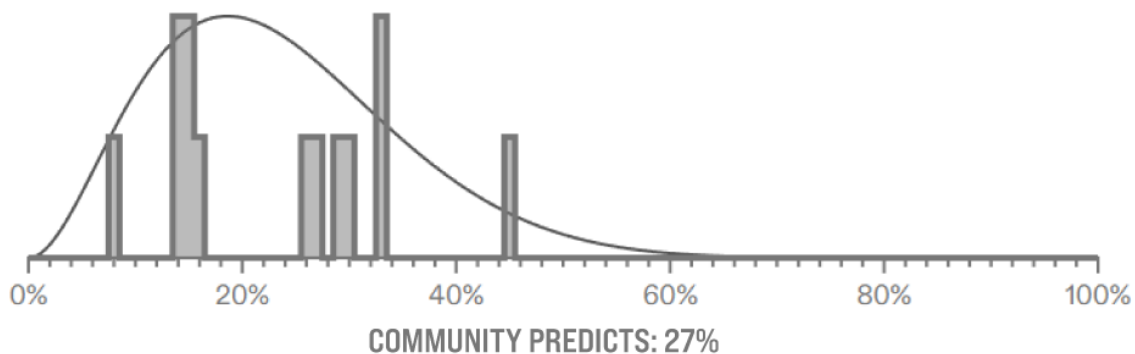
To assess the risks of escalation between the U.S. and China, the category *Measures of U.S.-China Tensions* contains two questions, one asking about Chinese aggression toward its neighbors and another about the probability of an incident between the U.S. and China. Not only would an air or maritime incident bring U.S. and Chinese forces into direct conflict, but Chinese escalation toward its neighbors poses a risk of this as well given the growing ties between the U.S. and Taiwan, the increased presence of the U.S. and its allies in the South China Sea, and Washington's overall sensitivity to aggressive actions from Beijing and vice versa.

⁶ The data in this report utilized to assess high impact publications consists only of peer reviewed research.

⁷ "AI scientific publications time series by country, from Scopus", OECD.AI Policy Observatory, <https://oecd.ai/en/data?selectedArea=ai-research&selectedVisualization=scientific-publications-time-series-by-country-2>

Will China attempt to undertake kinetic actions in contested territory before 2028?

As of September 19, 2022



In this context, kinetic actions in contested territory are defined as a military operation where Chinese troops enter the territory of another country or enter contested territory across a line of actual control and engage in hostilities resulting in 5 or more deaths total, combined from either side. Resolution is based on information from publicly available credible sources, and resolved by assessing the reliability of the information where reporting on deaths differs. **Forecasters estimate that there is a 27% chance that China will attempt to undertake kinetic actions in contested territory before 2028.** Forecasters construct their forecasts starting with a base rate from previous incidents that would likely satisfy the criteria, including the history of conflicts on the Indian border⁸ and the overall history of territorial battles waged by China.⁹ They then adjust from these base rates by assessing recent events and other evidence.

Several forecasters expect China is likely to try to take Taiwanese territory in the future, though are uncertain about this occurring within the next five years. Forecasters note that logistical challenges, the need for China to build its navy, and Chinese dependence on exports to Western countries may be significant barriers to an invasion in the short term. However, the Kinmen Islands, which are governed by Taiwan, may be vulnerable. Other forecasters note the poor outcomes for Russia thus far in its invasion of Ukraine may dampen China's outlook for its chances of success in Taiwan. However, some expect that, as with Russia's invasion of Ukraine, Chinese assessments of their chances of success may be skewed and overconfident.

⁸ "India-China border tensions: Key dates in decades-long conflict", Aljazeera, June 17, 2020, <https://www.aljazeera.com/news/2020/6/17/india-china-border-tensions-key-dates-in-decades-long-conflict>

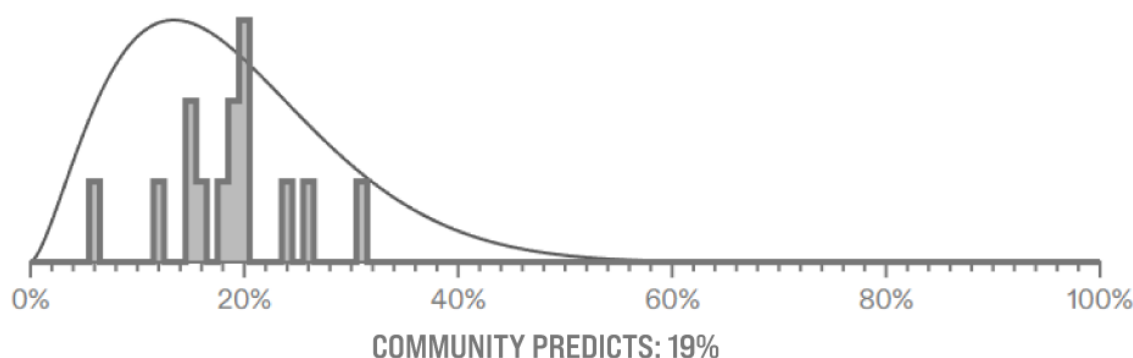
⁹ "List of wars involving the People's Republic of China", Wikipedia, Updated September 10, 2022, https://en.wikipedia.org/wiki/List_of_wars_involving_the_People%27s_Republic_of_China

With regard to the Indian border, forecasters note that recent skirmishes have generated risk, but also that recent signs of disengagement¹⁰ may make this less likely in the near term. It is also notable that forecasters extensively debated how to account for the Galwan Valley clash in 2020¹¹ due to the uncertainty in determining the conflict's exact location and whether China had indeed crossed the line of actual control, based on publicly available information. These uncertainties highlight some of the inherent difficulties in setting clear resolution criteria when predicting future events, as well as the challenge of establishing the facts on the ground using public information after events occur.

Regarding the probability of an air or maritime incident between U.S. and Chinese military forces prior to 2028, forecasters estimate a **19% chance** of this occurring. Forecasters note that activity in the South China Sea is increasing from both the U.S. and China, raising the risk of an incident. But the risk is counteracted by increased use of UAVs, which if shot down would not produce any military casualties as specified for this question. Civilian casualties also do not qualify towards resolution criteria, and the incident must involve an aircraft or naval vessel from either side. Forecasters also suggest that elevated tensions may cause the U.S. and China to take more care in the region to avoid accidental escalation, and that technological improvements in tracking and long-distance monitoring may enable more distance between military forces which might reduce the risk.

Will there be a military casualty caused by an air or maritime incident between U.S. and Chinese military forces before 2028?

As of September 19, 2022



¹⁰ Helen Davidson, "Indian and Chinese troops pull back from disputed Himalayan border area", The Guardian, September 9, 2022, <https://www.theguardian.com/world/2022/sep/09/indian-and-chinese-troops-pull-back-from-disputed-himalayan-border-area>

¹¹ Lauren Frayer, Laurel Wamsley, "20 Indian Troops Dead After Clashes With Chinese Soldiers Near Border", NPR, June 16, 2020, <https://www.npr.org/2020/06/16/877781736/3-indian-troops-dead-after-clashes-with-chinese-soldiers-near-border>

Nuclear Use

Finally, the *Nuclear Use* category contains two questions to quantify the likelihood of shifts in China's nuclear policy. Formal abandonment of "no first use" could constitute a significant change in nuclear policy, as could a nuclear detonation for testing or demonstration purposes.

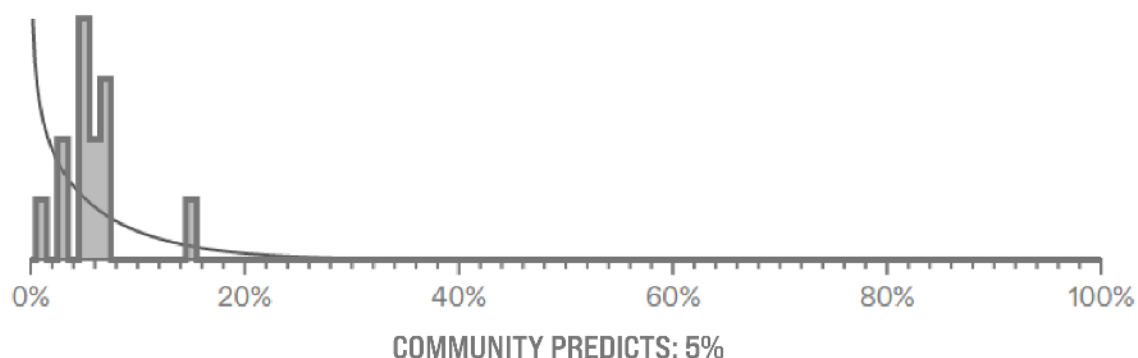
Forecasters estimate a 5% chance that China abandons "no first use" (NFU) by 2028.

Abandoning NFU is defined as a formal statement from the Chinese government or a Chinese government official that either explicitly revokes NFU or declares a policy that is clearly and unambiguously contradictory to NFU. The statement must be unretracted after one week in order to qualify. A nuclear first strike by China is interpreted as abandonment of NFU, unless there is strong evidence that the strike was not deliberate, presumed to be deliberate otherwise.¹²

The possibility that China shifts from this policy as a means of intimidation or to threaten countries who intend to aid Taiwan accounts for some of the affirmative probability assigned by forecasters. However, the need to rely on a nuclear first strike policy may indicate weakness, and adhering to NFU suggests confidence in their conventional warfare capabilities, giving China another reason to publicly adhere to it.

Will China formally abandon their "no first use" nuclear weapons policy before 2028?

As of September 19, 2022



Forecasters expect there may be limited use for nuclear testing today and in the future due to the availability of computer simulations, estimating an 8% chance of a Chinese nuclear test

¹² A deliberate strike was defined as one in which "the attacking nation decides to attack based on accurate information about the state of affairs", adopting the definition provided in Barrett et al. (2013). Anthony M. Barrett, Seth D. Baum & Kelly Hostetler (2013) Analyzing and Reducing the Risks of Inadvertent Nuclear War Between the United States and Russia, *Science & Global Security*, 21:2, 106-133, DOI: <https://doi.org/10.1080/08929882.2013.798984>

detonation by 2028. Recent reported activity at Lop Nur¹³ plays a significant role in adjusting forecasters' estimates upwards. General adherence to the Comprehensive Nuclear-Test-Ban Treaty among signatories and the lack of benefits to performing a test result in low forecasts for this question. However, forecasters generally believe the true probability of a test is higher, but the probability of publicly available knowledge of a test is low. Resolution for this question relies on the Arms Control Association to determine if there is sufficient evidence that a test occurred, but also notes that non-lethal "demonstrations" count as tests while lethal attacks do not count, and that a test need not be successful to qualify. If Metaculus has reason to believe the Arms Control Association does not follow the same criteria in its Nuclear Testing Tally¹⁴ then Metaculus will make final determination of whether or not an event qualifies.

What's Next?

Metaculus and IST plan to continue learning from these initial questions and expand on them to explore plausible security futures. We plan to generate a set of recommendations based on insights generated through the process. The partnership will also develop other future scenarios that integrate forecaster predictions, and will continue to field forecasts relevant to vulnerabilities and opportunities from AI and NC3 integration. Future questions will inform broader research into catastrophic risk, geopolitical stability, and digital security.

By using forecasting and subject matter expertise in assessing questions individually, categorically, and holistically, this partnership between IST's researchers and Metaculus's forecasters will provide an estimate of the risks associated with emerging technological phenomena and support more prepared decision-making as we navigate the most important problems in the coming decades.

¹³ Geoff Brumfiel, "A New Tunnel Is Spotted At A Chinese Nuclear Test Site," NPR, July 30, 2021, <https://www.npr.org/2021/07/30/1022209337/a-new-tunnel-is-spotted-at-a-chinese-nuclear-test-site>.

¹⁴ Daryl Kimball, "The Nuclear Testing Tally", Arms Control Association, August 2022, <https://www.armscontrol.org/factsheets/nucleartesttally>