

Expert Insights
PERSPECTIVE ON A TIMELY POLICY ISSUE

ABBY DOLL, KARL P. MUELLER

Exploring Instability Risks in the U.S.-China AI Rivalry

Breakwater Game Overview and Initial Observations

April 2026

For more information on this publication, visit **www.rand.org/t/PEA4703-1**.

About RAND

RAND is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. To learn more about RAND, visit www.rand.org.

Research Integrity

Our mission to help improve policy and decisionmaking through research and analysis is enabled through our core values of quality and objectivity and our unwavering commitment to the highest level of integrity and ethical behavior. To help ensure our research and analysis are rigorous, objective, and nonpartisan, we subject our research publications to a robust and exacting quality-assurance process; avoid both the appearance and reality of financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursue transparency in our research engagements through our commitment to the open publication of our research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. For more information, visit www.rand.org/about/research-integrity.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published by the RAND Corporation, Santa Monica, Calif.
© 2026 RAND Corporation
RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights

This publication and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited; linking directly to its webpage on rand.org is encouraged. Permission is required from RAND to reproduce, or reuse in another form, any of its research products for commercial purposes. For information on reprint and reuse permissions, visit www.rand.org/about/publishing/permissions.

PE-A4703-1

About This Paper

In this paper, we describe the design of *Breakwater*, a tabletop game created by a team of RAND researchers to examine competition dynamics and potential risks to international stability in the U.S.-China technological race toward the development of artificial general intelligence. We also summarize observations and preliminary hypotheses from the initial plays of the game. This paper is intended for readers who are concerned with tabletop simulation of issues related to artificial intelligence or designing games to inform policymaking about strategic competition, more generally, as well as for broader audiences who are interested in the national security implications of U.S.-China technological competition.

Center for the Geopolitics of Artificial General Intelligence

RAND Global and Emerging Risks is a division of RAND that delivers rigorous and objective public policy research on the most consequential challenges to civilization and global security. This work was undertaken by the division's Center for the Geopolitics of Artificial General Intelligence (AGI), which is committed to helping decisionmakers understand, anticipate, and prepare to navigate the national security and geopolitical implications of AGI. The center convenes leading technologists, strategists, economists, political scientists, and outside experts to consider the feasibility and effectiveness of prospective AGI-enabled capabilities; the domestic and international implications of their use; and the strategies and policies that governments, businesses, and civil society could adopt to respond to new realities. For more information, visit www.rand.org/geopolitics-of-agi.

Funding

This effort was independently initiated and conducted within the Center for the Geopolitics of Artificial General Intelligence using income from operations and gifts from RAND supporters, including philanthropic gifts made or recommended by DALHAP Investments Ltd., Ergo Impact, Founders Pledge, Charlottes och Fredriks Stiftelse, Good Ventures, Longview, and Coefficient Giving. RAND donors and grantors have no influence over research findings or recommendations.

Acknowledgments

Designing *Breakwater* has been a collaborative enterprise by a team that includes Julia Arnold, Hiwot Demelash, David Frelinger, Chad Heitzenrater, and Brian Jackson, in addition to the authors of this report. Joel Predd has unflaggingly supported and motivated the project, which has also benefited immeasurably from the contributions of many players and advisors from the larger Geopolitics of AGI community, both within and outside RAND. Most importantly, the success of

games, such as *Breakwater*, depends fundamentally on their players; we thank all of the colleagues and practitioners who have contributed their time, knowledge, and ideas to this effort. We are also indebted to Reid Pauly and Emily Yoder for reviewing this report and offering valuable suggestions to improve this final version.

Contents

About This Paper iii
Figures and Table vi

Exploring Instability Risks in the U.S.-China AI Rivalry: *Breakwater* Game Overview and Initial Observations 1
Breakwater Game Overview 1
Observations from Initial Gameplay 6
Emerging and Evolving Hypotheses 11
Next Steps 14

References 16
About the Authors 18

Figures and Table

Figures

Figure 1. U.S. and China National Gameboards 3
Figure 2. Strategic Competition and Disorder Gameboard..... 4

Table

Table 1. Participant Summary..... 7

Exploring Instability Risks in the U.S.-China AI Rivalry: *Breakwater* Game Overview and Initial Observations

The race to develop ever-more-advanced artificial intelligence (AI) is becoming increasingly prominent in the ongoing strategic rivalry between the United States and China. This intensifying technological competition raises the possibility that one or both of the contestants might seek to improve its prospects in the race—either to secure a lead or to keep the other from achieving artificial general intelligence (AGI) first—by attacking elements of its competitor's AI ecosystem using military force, cyber warfare, or other means beyond the bounds of normal peacetime statecraft.[1] Anticipating and managing these risks is important for both technology and national security policymakers, most of all because of the potential for such preventive actions to lead to armed conflict and potentially catastrophic escalation. As part of RAND's ongoing research into the geopolitical implications of AGI, our research team has developed and begun playing a game called *Breakwater*.[2] Specifically, *Breakwater* is designed to shed light on several interrelated questions: Under what conditions might competitors in the AI race resort to preventive actions, including cyberattacks or military force, to slow a rival's progress? How do beliefs about the trajectory and timeline of AI development shape strategic behavior and risk tolerance? What factors tend to stabilize or destabilize the competition as it unfolds?

This paper provides a brief overview of the game design and a summary of initial observations and hypotheses from early plays of the game.

Breakwater Game Overview

Breakwater is an analytic tabletop exercise (TTX) that explores how major-power competition over AI and AGI development could generate risks of international instability and conflict. In the game, Blue and Red player teams take on the roles of the U.S. and Chinese governments, respectively, and pursue strategies—either of their own invention or seeking to emulate current or proposed

[1] Karl P. Mueller, *Heeding the Risks of Geopolitical Instability in a Race to Artificial General Intelligence*, RAND Corporation, PE-A3691-12, July 2025. We are using the *AGI* label to refer to advanced AI that equals or exceeds human capabilities at performing a large share of important and complex tasks and is capable of a high degree of self-direction and self-improvement with limited human supervision.

[2] *Breakwater* is one of several games developed within this overall effort. For a description of a complementary game that focuses on investment strategies and international diplomacy in the pursuit of AGI, see David R. Frelinger, Jenny Oberholtzer, Gregory Smith, and Julia Arnold, *Insights from the Development and Play of Tsunami*, RAND Corporation, PE-A4240-1, October 2025.

national approaches to AI competition—to balance the following potentially competing objectives, prioritized based on their beliefs:

- Win the race to achieving AGI.
- Realize the benefits of AI advances in economic and military/intelligence sectors.
- Establish leadership and influence internationally in AI and AGI.
- Avoid overt conflict with the opponent.

There is no single victory condition in *Breakwater*. Instead, success is assessed holistically across these objectives during the postgame discussion, which reflects the reality that the AI competition does not have a clearly defined finish line and states must balance multiple, often competing, goals simultaneously. Events, such as paradigm shifts or geopolitical events, can also shift or change players' goalposts. This design choice means that players must weigh the risks of single-minded pursuit of any one objective, including raw technological acceleration, against the broader strategic consequences of neglecting the other objectives.

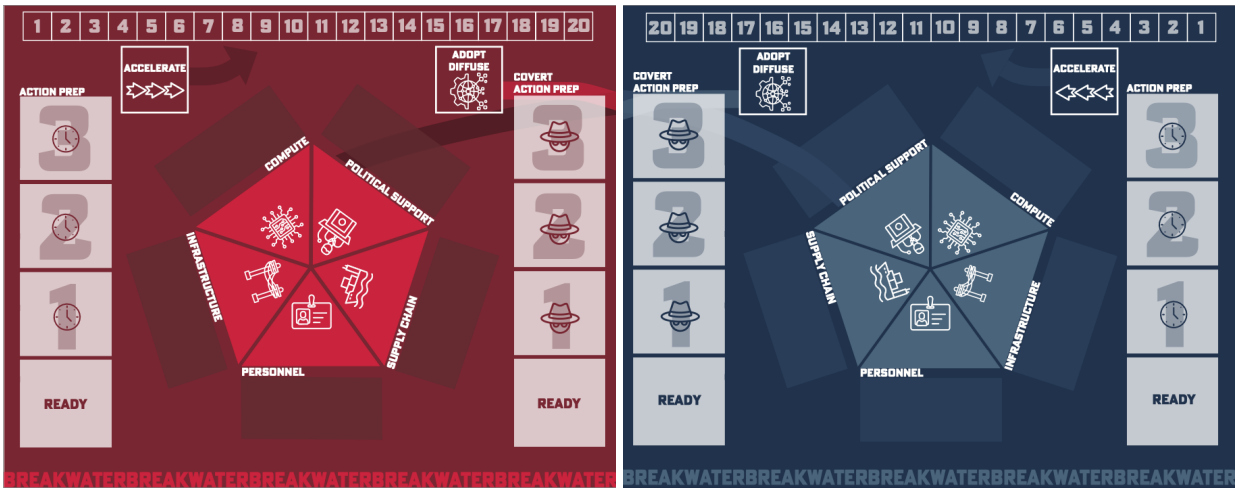
The Red and Blue teams compete over a series of game turns. During each turn, they make choices about allocating policy resources to accelerate their progress in foundational AI technology, promoting AI diffusion and adoption, taking preventive actions against their opponents' AI development, and protecting their AI ecosystems against potential hostile actions by their rival. The teams have parallel player boards that are used to track the overall progress of their AI models, actions that they take, and the status of their AI ecosystems (Figure 1). An AI Advancement Tracker shows the relative stage of development of each nation's (team's) AI models. Our baseline starting conditions posit that these stages represent levels of performance against prevailing benchmarks for hyperscaling multimodal or large language models (LLMs), although *Breakwater* abstracts the fine-grained details of AI development.[3]

We divide the states' AI ecosystems into five broad sectors: data and compute, supporting infrastructure (e.g., electricity generation and water for data centers), supply chain (e.g., chips, raw materials), personnel and talent, and political support for AI development and deployment. Limiting Factors, represented by red or orange markers, are placed within each sector either at the start of the game or as the result of opponent actions or event cards. *Limiting Factors* represent structural constraints—technical, political, or organizational—that hinder national or private company model development until their effects diminish or the affected side takes action to remove the impediments.[4] Players can also implement defenses or resilience measures to protect individual sectors from various types of offensive actions.

[3] As will be discussed later, we have the option of introducing a paradigm shift event during the game, which can rearrange players' positions of technological advantage and the types of limiting factors hampering their AI ecosystems.

[4] Red markers represent enduring limiting factors with effects that require major government intervention to alleviate their effects, such as lack of access to specialized chipmaking equipment, bureaucratic obstacles to infrastructure development required to support data centers, or brain drain effects of adversaries luring away key AI talent. Orange markers signify temporary limitations that can be expected to eventually be rectified through market responses, routine security and response measures, or decaying effects over time, such as repairable damage from enemy attacks or temporary disruptions to manufacturing supply chains.

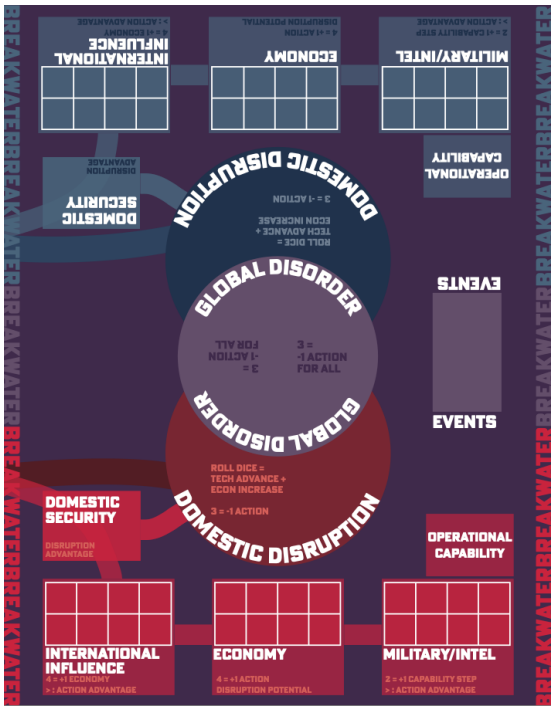
Figure 1. U.S. and China National Gameboards



A shared strategic competition board tracks how each team chooses to channel AI progress into concrete strategic advantages across key domains, as well as the effects of their actions in generating domestic or international disorder (Figure 2). As the teams' AI models advance along the tracker, they receive technological breakthrough markers. More markers appear at higher stages of development; players must choose how to invest them across the following four areas, which provide different types of benefits:

- **Economy:** Greater economic capacity increases the number of actions players may take each turn.
- **Military and intelligence:** Having a military/intelligence advantage over the rival power increases military operational effectiveness and improves the odds of success for military or covert actions.
- **International influence:** Superior international influence makes actions that depend on international support or cooperation more effective and generates additional economic momentum.
- **Domestic security:** Investing AI resources in domestic security increases the ability to prevent or suppress internal instability.

Figure 2. Strategic Competition and Disorder Gameboard



Breakthroughs appear more frequently as U.S. or Chinese AI advances, but rapid AI progress and adoption also increase the possibility of global disorder and domestic disruption, as do particularly aggressive or disruptive actions by the players. Reaching certain thresholds of domestic disruption imposes limits on that player's freedom of action, representing diversion of state resources and attention away from promoting AI toward crisis management; international disorder affects both sides. *Breakwater* includes three varieties of internal disruption effects: economic discontent, domestic social and political unrest, and AI safety and misalignment incidents. Disorder can arise because of aggressive or destabilizing player actions, from the broader effects of rapid AI progress and adoption occurring in the game, or through event cards that introduce exogenous shocks, such as geopolitical crises, AI safety incidents, or economic disruptions that are beyond either player's direct control. Player response options to deal with domestic disruption include both carrot and stick responses that have different chances of success and potential unintended consequences. Ameliorating global disorder requires cooperative action by the competitors.

Breakwater proceeds in six-month turns; each turn consists of simultaneous strategic moves followed by the adjudication of their consequences. Dice rolls and probability-based result tables and modifiers provide a baseline for adjudicating the success or failure of player actions and movement along the AI progress trackers. However, the game controllers have flexibility to make in-stride adjustments, and players are encouraged to draw on their expertise to debate outcomes or probabilities. The game's design includes several semistructured elements that allow for flexible initial starting conditions and trajectories of play. These elements include several possible AI-model development scenarios (which are initially hidden from the players) that shape each side's rate of

technological progress,[5] varying limiting factor starting conditions, and a library of offensive and defensive action cards that includes blank cards for players to add new actions to their menu of options in consultation with the gamemasters.[6] This adaptability is designed to serve one of the game's foundational purposes, which is to build on the underlying research and analysis by eliciting additional ideas from the players, and we emphasize to players that the game rules and components are intended to serve as the start of a discussion about AI competition techniques and strategies based on their own understanding and expertise. Thus, the game continues to evolve with repeated play as we iteratively refine the mechanics and game artifacts.

Within this structure, players have broad freedom to craft their strategy. At the start of the game, players are either provided an overall objective or strategic approach to guide their actions or asked to define their own objectives and describe their starting expectations about the technological trajectory toward AGI and their beliefs regarding their opponents' strategic intent. Each turn, they choose from among several categories of actions: acceleration or diffusion and adoption (represented by two-sided chips) or espionage, offensive operations, defensive bolstering, or managing disorder (represented by cards). Although both sides' options draw from these same broad categories, the specific cards available to each team differ in composition, difficulty level, and potential impact to reflect real-world asymmetries in U.S. and Chinese capabilities and strategic positions. Some cards may be played covertly, and many require one or more turns of preparation before they can be executed.

Players also choose how to invest breakthrough markers in the strategic advantage areas, take actions to remove limiting factors from their AI ecosystems, and may initiate military conflict.[7] They must also decide how to interpret and anticipate the opponent's behavior. These choices entail multiple trade-offs, including accelerating and protecting AI model development versus promoting broad adoption, prioritizing economic or military advantages, attacking the opponent's AI ecosystem versus protecting one's own, and balancing the risks of domestic disruption or global disorder against the anticipated benefits of pursuing faster gains. The gamemasters may also introduce event cards during the game to represent exogenous developments (such as geopolitical crises, technological disruptions, or domestic upheavals) that are outside the players' direct control but that may reshape the competitive landscape and force both teams to reassess their strategies. Overall, escalation considerations are central: Many actions that speed up progress or degrade the opponent's capabilities can also heighten instability by incentivizing the opponent to take more-aggressive actions.

[5] Current scenarios are a *plateau* (progress along the tracker has high probability of moving slowly), *graduated improvement* (progress along tracker slowly increases in probability as players reach different thresholds), and *rapid takeoff* (fast acceleration along tracker increases significantly once a certain threshold is crossed).

[6] The offensive and defensive cards span political, economic, cyber, kinetic, and espionage actions against the five AI ecosystem sectors. The action cards available to the teams are similar, but they include some options that are specific to China or the United States and some that are more likely to be successful for one side than the other, reflecting differences in national economies, military capabilities, and internal political systems.

[7] *Breakwater* has a rules module for resolving the effects of military engagements (including nuclear exchanges) should the players decide to initiate overt conflict. To date, players have not decided to do so.

Focus and Scope

Because *Breakwater* is designed to focus on stability dynamics and the potential for preventive actions in the AI competition, many aspects of the pursuit of AGI and the broader U.S.-China rivalry are abstracted or excluded from the game altogether. The game design does not include a detailed simulation of potential regional conflicts at the operational level, domestic and bureaucratic politics within the rival states, or the complex relationships between national governments and corporations that would also shape the environment in which the AI competition is taking place. However, we encourage players to include these dynamics in their deliberations during gameplay, and we have begun incorporating event cards to introduce exogenous factors that can impinge on the players' options and complicate their strategy development and execution. The game design abstracts technical details of AI architecture and specific legal and economic instruments, and it does not emphasize the potential for consequential actions by third-party state or nonstate actors. Our game system is also resource-agnostic besides limiting the number of actions the teams may play during a turn and including cost considerations in the difficulty levels assigned to particular actions. In general, given the objectives of the game and the desire to maximize opportunities for busy policy practitioners to participate, we seek to have players engaged in strategic choices without being bogged down by fine-grained policy planning and execution. However, we welcome discussion on these matters during adjudication or the structured post-game discussions (hotwashes).

Given the uncertainty surrounding AI's trajectory and its widely varied impacts, as well as the complexity of the broader AI ecosystem and competing player objectives, the game continues to evolve. Each new iteration refines the model to better capture emerging dynamics and ongoing strategic debates. Already, interacting with a wide variety of AI, economic, and military subject-matter experts during gameplays has led us to expand the menus of potential actions. It has also unearthed new questions.

Observations from Initial Gameplay

In this section, we summarize notable insights from twelve *Breakwater* playthroughs conducted between July and November 2025 (see Table 1). Players across the games comprised a broad variety of RAND subject-matter experts and a smaller number of external participants, including experts versed in China's strategic ambitions and approaches, former government officials, AI technologists, economists, and national security and foreign policy experts. Given the limited number of gameplays to date and the continuously evolving design,[8] these observations and hypotheses should be regarded as preliminary data and propositions, which we intend to explore further as *Breakwater* play continues and expands to include more and different players. Player beliefs and strategic dispositions are assessed through a combination of pregame questionnaires administered to participants, in-game observations and questions by the control team, and hotwashes in which players reflect on their reasoning and decisionmaking.

[8] Changes that have been implemented during this iterative design process have included additional action cards, differentiation between *carrots* and *sticks* in addressing domestic disruption, the addition of temporary limiting factors to reflect market forces or in-place response options, and military engagement rules (though no players have chosen to initiate overt conflict).

Table 1. Participant Summary

	Exercise											
Participants	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Current U.S. government staff	0	0	0	0	1	1	0	0	0	0	0	10
Former U.S. government officials (RAND)	0	0	4	0	0	0	0	1	0	0	0	0
Senior RAND or other analysts	4	4	1	4	1	1	3	1	3	2	2	4

Beliefs About Future AI/AGI Development Trajectories

Player expectations about the possibility, timing, and nature of AI/AGI takeoffs strongly shape strategic behavior during the game. Players are informed that the trajectory of AI progress could follow one of three potential paths (plateau, gradual stepwise improvement, or sudden takeoff), which makes rapid progress more or less likely at different points along the technology progress track. However, players do not know which scenario they are in. During the initial games discussed here, we always used a sudden takeoff scenario to explore how indicators that AGI may be imminent affect player approaches, but we varied the takeoff's timing and intensity.

When players believe that takeoff is imminent, they tend toward an all-in approach that emphasizes acceleration and removal of limiting factors at the expense of managing domestic disruption or cultivating international influence. Generally, players assume that the accrued risks are acceptable if the big payoff is around the corner, which will provide them with plentiful resources to address those issues. A sense of falling behind or a short lead can make escalatory or high-risk actions more attractive under this belief. However, an interesting pattern emerged: Even when players have been explicitly told that transformative AI could arrive any moment and they experience accelerated movement on the AI model tracker, players have not acted as if they expected a transformation immediately after the arrival of AGI so abrupt that the first state to possess AGI will have an insurmountable first-mover advantage. This dynamic appears to reflect beliefs that players brought into the game, but we also recognize that the game board design may also affect perceptions. Even though we state that the AGI threshold could be crossed at any point, the "long runway" of the AI tracker may make very early takeoff feel improbable.

When players expect at least a near-term trajectory of gradual AI improvement, they tend to distribute resources more evenly across investments in domestic stability, international influence, and incremental model advancement gains. In this mindset, international influence becomes a meaningful win condition rather than a secondary priority behind raw acceleration. Players also devote more attention to managing or seeking to avoid domestic disruption.

Players who remain uncertain or less committal about the trajectory of AI development tend to seek flexibility, maintaining moderate military buildup but avoiding premature escalation unless triggered by opponent actions. Defensive moves are framed as insurance, allowing players to pivot rapidly once the underlying AI trajectory becomes clearer. This uncertainty-based strategy often leads to the most-balanced initial action portfolios, with players hedging rather than betting on a single vision of the future.

Variations in Strategies and Objectives

Across our *Breakwater* playthroughs, U.S. and China players consistently choose to pursue different strategic objectives and implicitly run different races. U.S. players tend to view winning the technological frontier as central to success in the AI competition. Strategies pursued by U.S. players vary, including *monopolist acceleration*, in which players direct only selected supply-chain and cyber suppression efforts against China, and *managed escalation*, in which players hold offensive options in ready reserve to use in tit-for-tat responses against Chinese preventive actions. Some U.S. teams have followed preset strategies to explore their implications, such as seeking to emulate U.S. AI competition policy as reflected in official documents,[9] or implementing an AGI Rideout strategy that de-emphasizes racing to develop AGI first in favor of investing in AI ecosystem resilience and adaptive military capabilities to deter AI-enabled regional aggression that could escalate into major war.[10] In this strategy-testing mode, a central research focus is assessing the conditions and events that cause players to decide to deviate from their initial plans as the AI competition progresses.

China players, on the other hand, have generally pursued strategies that consider being second in the race as China's usual position with respect to the United States. These strategies range from *fast-following resilience*, in which players minimize choke-point exposures followed by aggressive acceleration or espionage, to assertively exploiting China's geo-economic influence and leverage, focusing on international influence and global market capture and holding rare earth restrictions and global talent recruitment as key levers. Players with China expertise regularly state that China is comfortable executing a fast-follower strategy so long as it secures early diffusion gains, increases international influence, and avoids falling dramatically behind the United States militarily. This asymmetry creates correspondingly asymmetric competitive dynamics: The United States often races for the frontier, while China focuses on international influence and quickly addressing domestic disorder when it arises. In combination, this intersection of strategies has not yet produced a highly unstable zero-sum race to AGI in which both sides consider reaching the finish line first to be imperative.

Neither have we observed player teams greatly overestimating how aggressive their opponent's strategy is, although the game provides opportunities for such misperception and suspicion—and for deliberate misdirection by the competitors regarding their intentions. Incorrectly imagining the rival to be pursuing a malign strategic agenda could lead to perceived incentives for preventive action even if both sides are defensively inclined. In future gameplay, we plan to explore further the conditions under which such spirals of hostility emerge in the game.[11]

Both sides have tended to initially allocate AI-derived advantages toward their economies rather than their militaries. In game terms, economic investments offer immediate benefits in the form of additional actions, whereas military advantages tend to feel speculative early on and only become

[9] White House, "National Security Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence," October 24, 2024; White House, *Winning the Race: America's AI Action Plan*, July 2025.

[10] David R. Frelinger and Karl P. Mueller, *The AGI Rideout Strategy for Reducing Strategic Risk and Promoting Stability in the Transition to Artificial General Intelligence*, RAND Corporation, PE-A4347-1, 2026.

[11] See Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics*, Vol. 30, No. 2, January 1978.

useful if an opponent becomes more offensively aggressive or a military capability gap is shifting rapidly and unfavorably. China players typically pursued military advantages more slowly; players have expressed caution about triggering U.S. responses and often wait for the United States to move first. Over time, however, when capability gaps widen or covert actions escalate, both sides eventually shift some of their focus toward military balance considerations.

U.S. players most often did not prioritize mitigating domestic disorder until it became a limiting factor on their available actions. When disorder arose, the U.S. players frequently acknowledged that the political costs (e.g., "Are we in an election year?") would likely shift attention and resources away from acceleration and efforts to suppress China's technological advances but were loath to divert many of their limited resources away from actions to promote progress in the AI race. On the other hand, China players repeatedly emphasized the importance of domestic stability to the Chinese Communist Party and rarely allowed domestic disruption to reach the threshold of interfering with their actions.[12]

Use of Cyber or Kinetic Preventive Actions

One of the main issues that *Breakwater* was designed to explore is how and under what circumstances players are more or less inclined to take preventive actions against their opponents with the goal of slowing adversary technological progress toward AGI. In theory, states should be strongly motivated to try to prevent rivals from achieving AGI if their leaders believe that doing so will dramatically increase a country's military and economic power. However, in recent history, preventive wars and similar actions have been very rare except for cases in which the target state has little ability to defend itself.[13] This rarity has a parallel in the nuclear domain, in which states that pursue nuclear weapons face an intense window of volatility, experiencing systematically elevated levels of armed conflict in the years surrounding acquisition.[14] Yet—even in the nuclear case, in which acquisition represents a dramatic and visible shift in military capability—preventive action has been rare against states that are capable of defending themselves. In *Breakwater*, players appear to perceive an even weaker case for more aggressive prevention: AI breakthroughs are not seen as producing the kind of immediate, decisive military advantage that would justify the risks of overt preventive action. The AI ecosystem's distributed and commercially redundant character further weakens the prevention calculus because it lacks the identifiable production bottlenecks, such as fissile material processing facilities, that make nuclear programs comparatively vulnerable to targeted strikes.

Consistent with this dynamic, players across our playthroughs have shown a strong and persistent preference for subkinetic tools over overt military strikes when seeking to impede their opponent's AI progress. With very few exceptions, neither U.S. nor China players have shown an inclination to launch kinetic attacks against data centers or other AI-related targets, even when players perceive

[12] As part of our assumptions regarding initial starting conditions, we assume that, based on its long-standing emphasis on maintaining strong internal control over its population, China has an initial advantage in tools for preventing and addressing domestic disorder.

[13] Zachary Burdette and Hiwot Demelash, *Evaluating the Risks of Preventive Attack in the Race for Advanced AI*, RAND Corporation, PE-A3691-13, September 2025; Hal Brands, *Don't Sweat the AGI Race: Why Concerns About Catastrophic Instability Are Overblown*, RAND Corporation, PE-A4188-1, September 2025.

[14] Vipin Narang, "Strategies of Nuclear Proliferation," *International Security*, Vol. 41, No. 3, Winter 2016–2017.

themselves to be falling behind. In contrast, many players have employed cyberattacks, in some cases repeatedly, and both sides have made extensive use of supply-chain pressure, espionage, talent-focused recruitment campaigns, and other measures that they consider to be relatively nonprovocative.

The preference for cyber over kinetic action does not appear to reflect a difference in expected material impact. In the game, neither category of action produces profound or enduring setbacks to adversary progress. Rather, this preference reflects a fundamental difference in how players categorize these tools. Players tend to view cyber operations as a normal and expected instrument of great-power competition; when targeted by a cyberattack, they rarely express surprise or outrage, often noting that this is something they anticipate a rival doing.[15] Kinetic strikes, by contrast, are viewed as crossing a threshold into overt conflict; such a strike carries escalation risks and political costs that players consistently judge to outweigh the limited disruption to AI progress that these attacks can achieve. The threat of a covert action (including cyberattacks) being detected by the adversary may also be deemed low risk because it still signals resolve and competitive intent while preserving space for the opponent to not respond in kind because of external pressures.[16]

Because the U.S. and China action decks differ in composition, difficulty levels, and potential impact (reflecting real-world asymmetries in capabilities), the two sides diverged in their offensive styles. China players consistently favored covert, incremental actions, leaning heavily on espionage, talent recruitment, probing of infrastructure and data centers, and information warfare to degrade political support for AI development. This reflected both the strategic constraints built into the game design and broader player perceptions: China lacked strong supply-chain retaliation options beyond rare earth export restrictions and robotics sector pressure, and players repeatedly expressed concern that overt disruption could trigger a U.S. response they could not counter. U.S. players typically pursued a more diversified and more overtly offensive approach, leveraging export controls, exploiting chokepoints in advanced manufacturing and chip design, and launching cyber operations against Chinese compute infrastructure.

Player-Specific Impacts on Gameplay

Teams of players controlling a side results in different behavior from when a single player controls China or the United States. Individual players tend to focus most deeply on specific sections of the policy space that match their expertise and tend to have more consistent strategy execution, while group dynamics for teams tend to lead to compromise strategies that result in spreading efforts and occasional pivots. This pattern is, of course, not unique to *Breakwater*, and it seems to parallel differences in national behavior between states led by individual strongmen and those with more pluralistic or collective decisionmaking processes.

The fact that previous sets of players did not resort to significant militarized or aggressive actions does not mean these outcomes are implausible. Additional game runs with different players who hold more hawkish or risk-acceptant beliefs will help confirm or revise this observation. In addition, we will

[15] Erica D. Borghard and Shawn W. Lonergan, "Cyber Operations as Imperfect Tools of Escalation," *Strategic Studies Quarterly*, Vol. 13, No. 3, Fall 2019.

[16] Austin Carson and Keren Yarhi-Milo, "Covert Communication: The Intelligibility and Credibility of Signaling in Secret," *Security Studies*, Vol. 26, No. 1, 2017.

explore how different geopolitical event cards may exacerbate or upend players' beliefs about the race and their opponents, introducing exogenous instability dynamics that could spill into the AI competition.

Emerging and Evolving Hypotheses

At this relatively early stage of *Breakwater*'s development life cycle, it would be premature to draw conclusions from patterns of player behavior in the games that have been conducted. However, observations to date suggest some hypotheses that will figure prominently among priorities for testing in future iterations of the game as we seek to gain insight into the geopolitical dynamics of the AI race.

The Longer AGI Takes to Manifest, the Less Destabilizing Its Arrival Is Likely to Be

A longer timeline for evidence of AGI's emergence may lead to a transition that is less destabilizing than a short timeline because both sides will have had more opportunity to adapt to steadily improving AI capabilities and invest in strategic advantages while observing and counterbalancing their opponents' moves. Generally, our game playthroughs introduced takeoff dynamics after the third or fourth turn. To further test this hypothesis, the game should include more iterations in which AGI-level advances arrive quickly and with minimal warning; these runs would reveal which stabilizing strategies still function when players have little time to adjust, which ones break down, and which new incentives or escalatory pressures emerge only under fast-takeoff conditions. It will also help us refine and expand our understanding of stability in a world after AGI is achieved, as opposed to during the race to achieve it.

Perceived Urgency in the AI Race Is Driven More Powerfully by Tangible AI Investments Providing Applied Strategic Advantages Than by Hypothetical Predictions

Across the playthroughs, players often had stronger reactions to shifts in relative strategic advantages (economic, military/intelligence, international influence) than to modest shifts of position on the AI progress tracker. Players tended to react this way because the strategic advantages produced tangible effects in the game. Achieving certain levels of economic expansion allowed for more actions per turn, and relative advantages in international influence and military/intelligence capability gave advantages or disadvantages for related action cards. If one side gained a significant lead in one area of competition, particularly in the economic or military/intelligence areas, opponents soon redirected resources to acceleration or shifted how they prioritized their strategic advantage investments to regain lost ground. Together, these reactions support the hypothesis that perceived urgency in the race to AGI is driven less by theoretical model sophistication estimates than by how AI-enabled capabilities translate into real shifts in national power. This suggests that AGI-race stability may depend less on relative model or frontier progress and more on how states choose to invest, signal, or conceal applied

advantages. To further test this hypothesis, the game could explore both specificity and uncertainty about what the currently abstract "strategic advantage" markers represent, allowing players to describe their priorities in that sector or attempt to conceal their investments.

Kinetic or Cyberattacks Likely Need to Be Systemic and Sustained to Significantly Impede Adversary AI Progress

We based the action card ratings and adjudication system on our research that has, thus far, identified no credible "silver bullet" kinetic military or cyber options that are capable of decisively slowing or stopping a rival's AI development while being sufficiently limited to avoid triggering substantial escalation risks (and none that appear capable of reversing a rival's AI development progress). Both domains offer tools that create friction, impose temporary delays, or disrupt specific components of an AI ecosystem, but none can reliably generate long-duration setbacks at limited political or military cost. Military strikes on data centers or fabrication facilities require large numbers of precision munitions, expose the forces launching them to attack or retaliation, and would almost certainly be viewed by adversaries as strategic attacks. We postulate the same to be true for lower-level sabotage attacks, which would need to be widespread and sustained, all without attribution. Generally, work can be relocated, capacity is geographically distributed, components are replaceable, and firms can surge procurement; we have yet to identify truly scarce, long-lead-time bottlenecks that targeted, one-and-done kinetic attacks against small numbers of targets could exploit for major effect.

Cyber operations have their own hard limits. Even sophisticated actions require long preparation, depend on fragile access vectors, and often produce only temporary or localized effects because of redundancy, software backups, and rapid reconstitution in modern AI infrastructure. Cyber operations can corrupt data, disrupt compute pipelines, or degrade training runs, but these effects are usually reversible within days to weeks once detected and entail uncertainty about whether the desired impact will land at the right time—or even the right place.

These findings directly shaped how we designed our action cards and adjudication probabilities and modifiers. The difficulty levels of these aspects of the game reflect our observed real-world challenges of access, planning, and execution. Their impact levels reflect the limited and context-dependent nature of the damage that they can individually cause, and their duration reflects our assumptions on how quickly modern AI ecosystems can rebuild or route around disruptions. As a result, attempts by players to meaningfully impede an opponent's AI progress through kinetic attacks or cyberattacks are rarely decisive, often risky, and almost always carry significant escalation hazards, which seriously limits the appeal of such actions to players in the game. This pattern is consistent with the past infrequency of overt preventive actions against powerful adversaries in the real world. We will continue to refine our understanding of attack impacts and response abilities through further research, future iterations with additional subject-matter experts, and gameplays at varying levels of classification.

Strategies to Stabilize the AGI Race Through Preventive Attack Threats Will Lack Credibility

The practical obstacles to disabling adversary AI development efforts through limited uses of force severely undermine the plausible feasibility of such strategies as the Mutual Assured AI Malfunction (MAIM) concept by Hendrycks, Schmidt, and Wang, which proposes using threats of preventive strikes against data centers and related targets to deter states from developing dangerous AGI.[17] Even a very optimistic assessment of the military requirements for stopping a rival's AGI development in its tracks should anticipate the need for a large-scale use of force against many high-value targets, which history suggests Chinese or U.S. national leaders would be very unlikely to undertake unless a major war were already underway.[18] However, players' discussions and decisions in *Breakwater* games to date have shown the possibility that these leaders may be nearly as reluctant to launch even very limited kinetic attacks against their powerful, nuclear-armed rivals because of the potential for such strikes to lead to further escalation. If an adversary believes this possibility to be true, whether correctly or not, the deterrent value of such strategies would become even more doubtful.

An AGI Paradigm Shift Could Create Significant Instability by Upending Strategies and Familiar Benchmarks

In one game, we introduced a "paradigm shift" event card, which represented a sudden Chinese breakthrough that disrupts the familiar hyperscaling LLM trajectory and realigns the structure of advantage in the AI race.[19] Under this scenario, years of optimization for one technological path are abruptly rendered less relevant, forcing both sides into a disruptive and uncertain transition.

We chose to explore this event to test players' reactions to a plausible real-world development—U.S. big bets being placed on the wrong race. The United States may continue to optimize for a frontier foundation-model strategy, in which scaling large generative models is the core engine and everything else (agents, tool retrieval, etc.) is built around it.[20] If China's research ecosystem devotes greater investment to embodied learning, world-model architectures, algorithmic-efficiency breakthroughs, or other non-hyperscaling approaches, these other bets could allow China to leap forward in the competition to develop AGI relative to the United States, despite China trailing in

---

[17] Dan Hendrycks, Eric Schmidt, and Alexandr Wang, "Superintelligence Strategy: Expert Version," *SuperIntelligence—Robotics—Safety & Alignment*, Vol. 2, No. 1, March 2025. The authors argue for the merits of a regime in which the major powers all deter each other from building AGI through fear that their AI programs will be destroyed if they attempt to cross the threshold to AGI, which the authors argue would pose an existential threat to the world in any hands.

[18] Regarding the MAIM strategy's questionable assumptions, see Iskander Rehman, Karl P. Mueller, and Michael J. Mazarr, "Seeking Stability in the Competition for AI Advantage," commentary, RAND Corporation, March 13, 2025.

[19] In game terms, the event dramatically reshuffles positions on the AI technology tracker: The United States is displaced by China and pauses as it searches domestically for institutions capable of adopting to the new approach, resulting in temporary personnel and talent limiting factors. China loses its compute impediments because the new paradigm relies less on GPU-driven hyperscaling and more on hardware categories that impose fewer constraints.

[20] Currently, the most visible U.S.-based frontier efforts with the largest resource commitments are focused on foundation models with transformer-style scaling. See, for example, William Marcellino, Lav Varshney, Anton Shenk, Nicolas M. Robles, and Benjamin Boudreaux, *Charting Multiple Courses to Artificial General Intelligence*, RAND Corporation, PE-A3691-1, April 2025.

LLM capabilities. In such a world, U.S. AI developers would need to rapidly pivot their resourcing, architecture, talent recruitment, and development strategies.

Player reactions during the playthrough suggested that this could introduce a strategic upheaval. The China players had, until that point, been pursuing a diffusion-oriented, fast-follower strategy to avoid provoking the United States. When they were vaulted into the lead, they became anxious because they had not invested heavily in defense or resilience and felt exposed and at risk of potentially desperate or fear-based U.S. offensive actions. Meanwhile, the U.S. players reacted with uncertainty and confusion because their established playbooks for countering a technologically inferior China no longer applied and their familiar metrics for judging progress no longer mapped cleanly onto the new AI competition landscape. Both teams experienced an immediate loss of strategic clarity, highlighting how destabilizing an overturn of long-standing assumptions about the nature of AI's progress could be.

As illustrated by the fervor following DeepSeek's release of high-performance models trained with far less compute and dramatically lower cost, a real-world paradigm shift could lead to destabilizing beliefs and behaviors.[21] By revealing that breakthroughs could emerge from efficiency gains rather than hyperscaling, DeepSeek exposed potential vulnerabilities in U.S. assumptions about which inputs drive AI leadership and which kinds of constraints are actually load-bearing. Events such as these undermine shared expectations about how to measure progress, who is strategically exposed, and which development portfolios are more resilient. Because both states may optimize their strategies around assumed technological pathways, an expected leap forward by a trailing state would force rapid, high-stakes adaptation under uncertainty. This could create conditions in which domestic disorder, misread signals, and reactive militarization become more likely as each side struggles to reinterpret the strategic landscape.

Additional research into potential paradigm shift scenarios will allow further exploration of this hypothesis. For example, we may create events involving different types of shifts in the relative importance of AI ecosystem components, as well as exploring the differences between gradual shifts and more sudden and unexpected ones. It would also be worth exploring a variant event in which the paradigm shift disproportionately harms the follower, rather than the leader, to examine the dynamics of a trailing competitor abruptly falling much farther behind in the competition, which could happen for multiple reasons beyond paradigm shifts.

Next Steps

Breakwater requires teams to navigate simultaneous technological competition, geopolitical maneuvering, and stability management under a high degree of uncertainty. Players confront bounded resources, imperfect information, escalating risks, and shifting incentives, which mirrors the strategic tensions that the game is designed to illuminate. Next steps for this effort include running more scenarios that examine post-AGI dynamics, as well as the earlier-stage race conditions that have dominated initial playthroughs; systematically testing competing strategies across repeated games; and

[21] Eduardo Baptista, "Explainer: What Is DeepSeek and Why Is It Disrupting the AI Sector?" Reuters, January 28, 2025; Eva Dou, Aaron Gregg, Cat Zakrzewski, Nitasha Tiku, and Shannon Najmabadi, "Trump Calls China's DeepSeek AI App a 'Wake-Up Call' After Tech Stocks Slide," *Washington Post*, January 27, 2025.

expanding the variety of participants to include players with a wider variety of expertise and policymaking experience at different levels and from different professional communities, as well as a more-diverse variety of perspectives on the likely trajectories and timelines of advanced AI development and of the U.S.-China strategic rivalry.[22] The next phase of development for the *Breakwater* design will emphasize fleshing out the event card system to introduce exogenous shocks that disrupt expectations and reshape competitive behavior. In future iterations of the design, we intend to explore further representing the role of private companies in U.S. decisionmaking about AI and deepening the modeling of China's internal political stability to more accurately capture how domestic incentives influence its AI strategy.

[22] Priority areas for player engagement include representatives from the AI and technology sector, current staff within key government agencies, and members of other AI research organizations.

References

Baptista, Eduardo, "Explainer: What Is DeepSeek and Why Is It Disrupting the AI Sector?" Reuters, January 28, 2025.

Borghard, Erica D., and Shawn W. Lonergan, "Cyber Operations as Imperfect Tools of Escalation," *Strategic Studies Quarterly*, Vol. 13, No. 3, Fall 2019.

Brands, Hal, *Don't Sweat the AGI Race: Why Concerns About Catastrophic Instability Are Overblown*, RAND Corporation, PE-A4188-1, September 2025. As of April 2, 2026: https://www.rand.org/pubs/perspectives/PEA4188-1.html

Burdette, Zachary, and Hiwot Demelash, *Evaluating the Risks of Preventive Attack in the Race for Advanced AI*, RAND Corporation, PE-A3691-13, September 2025. As of April 2, 2026: https://www.rand.org/pubs/perspectives/PEA3691-13.html

Carson, Austin, and Keren Yarhi-Milo, "Covert Communication: The Intelligibility and Credibility of Signaling in Secret," *Security Studies*, Vol. 26, No. 1, 2017.

Dou, Eva, Aaron Gregg, Cat Zakrzewski, Nitasha Tiku, and Shannon Najmabadi, "Trump Calls China's DeepSeek AI App a 'Wake-Up Call' After Tech Stocks Slide," *Washington Post*, January 27, 2025.

Frelinger, David R., and Karl P. Mueller, *The AGI Rideout Strategy for Reducing Strategic Risk and Promoting Stability in the Transition to Artificial General Intelligence*, RAND Corporation, PE-A4347-1, April 2026. As of April 27, 2026: https://www.rand.org/pubs/perspectives/PEA4347-1.html

Frelinger, David R., Jenny Oberholtzer, Gregory Smith, and Julia Arnold, *Insights from the Development and Play of Tsunami*, RAND Corporation, PE-A4240-1, October 2025. As of April 9, 2026: https://www.rand.org/pubs/perspectives/PEA4240-1.html

Hendrycks, Dan, Eric Schmidt, and Alexandr Wang, "Superintelligence Strategy: Expert Version," *SuperIntelligence—Robotics—Safety & Alignment*, Vol. 2, No. 1, March 2025.

Jervis, Robert, "Cooperation Under the Security Dilemma," *World Politics*, Vol. 30, No. 2, January 1978.

Marcellino, William, Lav Varshney, Anton Shenk, Nicolas M. Robles, and Benjamin Boudreaux, *Charting Multiple Courses to Artificial General Intelligence*, RAND Corporation, PE-A3691-1, April 2025. As of April 9, 2026: https://www.rand.org/pubs/perspectives/PEA3691-1.html

Mueller, Karl P., *Heeding the Risks of Geopolitical Instability in a Race to Artificial General Intelligence*, RAND Corporation, PE-A3691-12, July 2025. As of April 9, 2026: https://www.rand.org/pubs/perspectives/PEA3691-12.html

Narang, Vipin, "Strategies of Nuclear Proliferation," *International Security*, Vol. 41, No. 3, Winter 2016–2017.

Rehman, Iskander, Karl P. Mueller, and Michael J. Mazarr, "Seeking Stability in the Competition for AI Advantage," commentary, RAND Corporation, March 13, 2025. As of September 18, 2025: https://www.rand.org/pubs/commentary/2025/03/seeking-stability-in-the-competition-for-ai-advantage.html

White House, "National Security Memorandum on Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence," October 24, 2024.

White House, *Winning the Race: America's AI Action Plan*, July 2025.

About the Authors

Abby Doll is a senior political scientist at RAND. Her work focuses on military capabilities assessment and military strategy. She holds a Ph.D. in war studies.

Karl P. Mueller is a senior political scientist at RAND. His research focuses on deterrence and coercion, airpower, escalation, crisis stability, and wargaming. He holds a Ph.D. in politics.