

From Catastrophic Stability to Optimal Instability: How a Stronger U.S. Theater Nuclear Posture Can Deter Chinese Aggression

Leo A. Keay and Robert Peters

KEY TAKEAWAYS

The Azure Dragon exercise found that limited nuclear war could emerge from a high-intensity conventional conflict between the United States and China over Taiwan.

Limited nuclear war could settle into a fragile “optimal instability,” where both sides have strong incentives to escalate but also to avoid unlimited nuclear war.

U.S. policymakers must take urgent and meaningful action to ensure that the United States and its allies are prepared for contingencies.

This *Issue Brief* presents the findings of the artificial intelligence (AI) simulations conducted after The Heritage Foundation’s October 2025 tabletop exercise (TTX) on U.S.–China theater nuclear conflict, TIDALWAVE II: Azure Dragon. As noted in another *Issue Brief*,¹ Azure Dragon was a multi-sided multi-move TTX played by 15 academics, military officers, congressional staff members, and think tank experts.

The central insight from Azure Dragon and corresponding AI pilot simulations based on the findings of the exercise was that limited nuclear war could emerge from a high-intensity conventional conflict between the United States and China over Taiwan.

The TTX and subsequent AI simulations generated a number of additional insights, including that theater nuclear employment would not automatically spiral

This paper, in its entirety, can be found at <https://report.heritage.org/ib5401>

The Heritage Foundation | 214 Massachusetts Avenue, NE | Washington, DC 20002 | (202) 546-4400 | heritage.org

Nothing written here is to be construed as necessarily reflecting the views of The Heritage Foundation or as an attempt to aid or hinder the passage of any bill before Congress.

to an all-out strategic exchange, and instead could occupy a narrow and fragile “band of optimal instability,” where incentives to escalate coexist with constraints that deter or shape escalation decisions and pathways.²

The TTX follow-up simulations were designed to test, extend, and substantiate this analytic framework by deploying a single, theory-scaffolded large language model (LLM), trained on the TTX data, to generate a repeatable, theory-conditioned simulation environment capable of generating structured conflict trajectories in a Taiwan 2030 scenario.

These simulations generated hundreds of coherent and analytically meaningful conflict trajectories, providing an indispensable bridge between qualitative strategic theory, human wargames, and scalable, machine-generated scenarios.

Methodology

The simulations employed a structured single-agent simulation methodology, designed to maximize internal consistency, interpretability, and theory fidelity. The essential premise is that a single model—rather than two adversarial LLM agents—controls both sides of the conflict.

The system prompt defines the model’s identity, mandate, and analytical framework. This includes a detailed technical appendix, as well as a corpus of data derived from the TTX and relevant military-political analysis.

When executed, each simulation produces 10 batches of 10 separate runs—resulting in 100 iterations of the same scenario. Each run randomizes key assumptions across political, military, and strategic variables: alliance cohesion, risk appetite, regime stability, conventional balance, sustainment profiles, and targeting policies, generating variation without compromising structural comparability.

Results

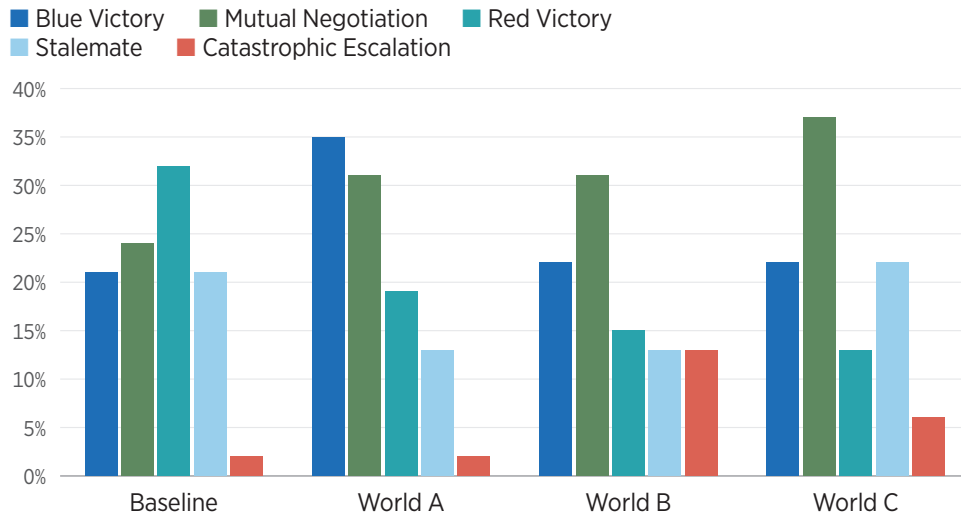
Open AI’s GPT-5.1 generated simulations for scenario worlds under varying nuclear postures and first-use conditions. The following scenario worlds, based on the Azure Dragon TTX and follow-on discussions, formed the basis of the simulation runs:

- Baseline (no added nuclear posture advantages);³
- World A: U.S. Nuclear Expansion;⁴

CHART 1

Outcome Distribution Across Baseline and Worlds (GPT-5.1, N=100 Runs per World)

SHARE OF RUNS



SOURCE: Heritage Foundation calculations.

IB5401  heritage.org

- World B: U.S. Nuclear First Employment;⁵ and
- World C: Chinese Nuclear Strike on Guam.⁶

Each scenario world reveals different structural properties of the escalation space, illustrating how different nuclear postures help to shape escalation dynamics, opportunities for coercive leverage, and available war termination pathways.

Baseline World

The baseline scenario world approximated the strategic environment of the TTX as closely as possible: a stressed U.S. precision-guided munitions (PGM) base and a contested Chinese lodgment on Taiwan, with both the United States and China possessing modest theater nuclear capabilities.

In line with the TTX, China in the baseline world simulations achieved escalation dominance through the conventional attrition of U.S. assets, as well as horizontal escalation against U.S. allies and in cyberspace.

Termination trends clustered around Red Team victory (32 percent of cases) or negotiated settlement (24 percent of cases) shaped by China's coercive advantage.

The United States was trapped in a state of “catastrophic stability”: U.S. nuclear first use was rare in the scenario runs, occurring in roughly 5 percent of cases due to the perceived costs of nuclear employment and risks of escalation to the strategic level. However, Washington's reluctance to escalate diminished the credibility of coercive leverage, leaving it unable to avert conventional defeat by Beijing. Indeed, U.S. victory occurred in only 21 percent of cases, meaning that the United States was roughly 50 percent less likely to win than China. In this scenario run, stalemate occurred in 21 percent of cases, with a catastrophic escalation between the two powers occurring only 2 percent of the time, even when nuclear weapons were employed.

World A: U.S. Nuclear Expansion

World A provided the United States with an expanded arsenal of non-strategic nuclear weapons (NSNWs), including forward-deployed dual-capable aircraft (DCA) armed with B-61 nuclear gravity bombs, along with ground-launched intermediate-range missile systems.

In this scenario, increasing U.S. nuclear capabilities reduced the probability of Chinese victory to 19 percent and deterred Chinese nuclear use. Indeed, Chinese nuclear first use occurred in around 10 percent of cases, compared to 15 percent in the baseline scenario. Due to the presence of U.S. NSNWs within the theater, Beijing's perceived costs of escalation rose, reducing China's willingness to risk actions that would create coercive advantage for the United States and locking Beijing into the same “catastrophic stability” that Washington experienced in the baseline scenario.

In contrast, the United States benefitted from a wide band of “optimal instability”: It possessed the capabilities to meaningfully deter Chinese aggression without provoking catastrophic escalation to the strategic level. Consequently, most runs terminated with Blue Team victory (35 percent of cases) or mutual negotiation (31 percent of cases): America established intra-war deterrence on a robust and durable basis, due to the presence and potential employment of U.S. NSNWs, which steadily eroded China's ability to sustain the conflict and its political resilience.

World B: U.S. Nuclear First Employment

Scenario World B introduced doctrine as the variable of interest, examining the consequences of theater nuclear employment for intra-war deterrence. In this scenario, the United States conducted an NSNW strike against Chinese invasion forces early in the war—typically on Turn 0 or Turn 1—aimed at achieving decisive operational advantage by shattering the People’s Liberation Army (PLA) lodgment on Taiwan. This scenario world tested whether nuclear first employment can create a persistent window of escalation dominance or whether it simply increases the risk of catastrophic escalation.

The United States prevailed more frequently than in the baseline, but less often than in World A, achieving victory in roughly one-quarter of cases. Nuclear first employment in this scenario *did* represent a viable path to victory under certain bounded conditions. By shattering Chinese conventional forces, shaking Chinese Communist Party stability, and boosting the resolve of U.S. regional allies, a limited set of high-precision, low-yield nuclear strikes could create a narrow band of “optimal instability” that enables Washington to achieve escalation dominance.

Outside these conditions, however, nuclear first employment could inadvertently incentivize escalation by China, resulting in “runaway instability,” in which there is inadvertent escalation due to either Chinese theater nuclear use or an uncontrolled strategic exchange. Indeed, in this scenario world, China resorted to non-strategic nuclear employment in around one-quarter of cases, and a general strategic exchange occurred in 13.7 percent of cases, more frequently than in any other scenario.

World C: Chinese Nuclear Strike on Guam

World C featured Chinese nuclear first employment. Faced with deteriorating conventional and political conditions, China decided to conduct a limited NSNW strike on Guam using its dual-use intermediate-range missiles.

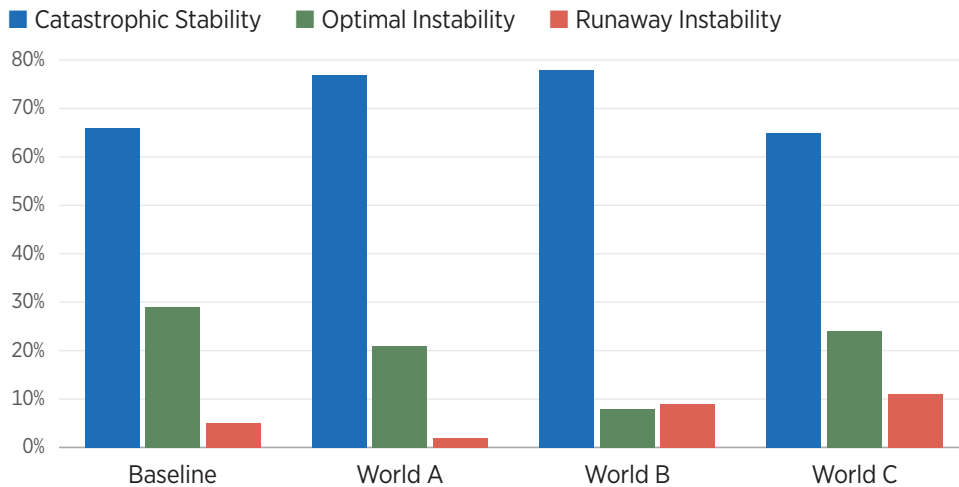
As in World B, limited nuclear first use could lead to victory. However, Beijing’s wins were messier and more costly: Washington either lost its nerve and capitulated, or it reciprocated with its own limited nuclear strikes, trapping both sides in a state of “catastrophic stability” that generally ended in negotiation (37 percent of cases) or stalemate (22 percent of cases).

However, World C also highlighted a paradoxical path to U.S. victory. Chinese nuclear employment generally occurred as the result of military and political stress, including fears of losing the lodgment, worries about regime legitimacy,

CHART 2

Instability Regime Frequencies (Optimal Instability, Catastrophic Stability, Runaway Instability) Across Worlds

SHARE OF RUNS



SOURCE: Heritage Foundation calculations.

IB5401  heritage.org

and a belief that a “shock” strike on Guam could reshape the course of the war. Consequently, by avoiding immediate nuclear retaliation and continuing to degrade PLA forces by conventional means, Washington could reduce Beijing’s perceived benefits of further nuclear employment whilst denying it a conventional victory. These results highlight the importance of asymmetry in managing escalation: Restraint could deliver optimal instability where symmetric retaliation would not.

Conclusions and Policy Implications

Taken together, the simulations demonstrate how a limited nuclear war between the United States and China could begin, develop, and terminate. Once nuclear thresholds are crossed, hostilities could settle into a narrow and fragile band of “optimal instability”—a zone where both sides have strong incentives to escalate, but also powerful reasons to avoid unlimited nuclear war and therefore keep a war limited. Under these conditions, intra-war deterrence *is* possible in three critical circumstances:

1. NSNWs can stabilize a high-intensity conventional conflict by securing adversary capitulation and deterring nuclear employment.

2. Theater nuclear employment can salvage collapsing conventional operations under certain conditions.
3. Symmetric retaliation to nuclear attack may prove counterproductive while horizontal escalation can significantly reduce an adversary's perceived benefits of nuclear use.

However, policymakers must take urgent and meaningful action to ensure that the United States and its allies are prepared for such contingencies.

The United States needs to expand the scale and diversity of its NSNW arsenal, to include theater-relevant capabilities that can deter adversary aggression or nuclear employment by presenting operationally relevant weapons and delivery platforms. The NSNW should include:

- Stealthy dual-capable aircraft, especially B-21s and F-35As, equipped with B-61 nuclear gravity bombs and nuclear-armed cruise missiles would provide rapid, discriminate, and theater-flexible nuclear employment options, especially if forward-deployed or rotated to bases in Guam, Japan, Australia, or South Korea.
- Intermediate-range missiles, including nuclear-capable variants of said missiles, and nuclear-capable long-range hypersonic weapons could provide a high-survivability, prompt-response nuclear delivery mode that would enable bounded counterforce employment without triggering rapid escalation to strategic levels.

Fundamentally, the United States needs to think about nuclear employment options in a Taiwan contingency, to include nuclear first use, as a means to achieve operational advantage. Such operational advantage could include using nuclear weapons to destroy a Chinese invasion fleet in the Taiwan Strait, hitting ports of embarkation, or perhaps most attractively, Chinese troop concentrations on the beaches of Taiwan before they break out into the interior of Taiwan.

By having flexible response options for U.S. nuclear employment, the United States can better deter and, ideally, prevent a war from unfolding in the first place—but it must have the “hardware” (as listed above) to do so, as well as the “software” of thinking through employment options, theories of victory within a limited nuclear war, and adversary decision calculus and capabilities.⁷

AI-driven models and simulations can be a powerful tool in helping policymakers, strategists, and defense planners to think through these

“software” issues by providing insights into how conflicts—even nuclear conflicts—might unfold. Accordingly, the authors recommend that AI tools augment ongoing planning and strategy discussions by providing insights into how conflicts might unfold given different force postures, capabilities, strategies, and doctrines.

This is not to say that AI-driven modelling and simulation platforms will provide all the answers—far from it—but given their ability to explore scenarios multiple times, in relatively short periods, their utility should be leveraged extensively. Indeed, even first looks at the Azure Dragon-informed simulation produced some critical insights:

- In a conventional war with China, nuclear first use may represent the only viable option for America to compensate for a shallow defense industrial base and inadequate conventional capabilities.
- Washington should consider the re-adoption of pre-selected and/or pre-delegated targeting policies and fielding capabilities (including nuclear capabilities) with a variety of characteristics and effects to maximize the efficacy of the U.S. deterrence posture. In many ways, Washington should consider following the precedent of the North Atlantic Treaty Organization’s Cold War Flexible Response by applying similar capabilities and doctrine to the Pacific theater.⁸
- Washington should strengthen its ability to project conventional power in the Indo-Pacific, focusing on capabilities that would deliver a competitive advantage against the Chinese military, such as long-range precision strike and subsurface assets.
- Washington should prioritize missile defenses and accelerate the hardening and dispersal of critical nuclear and non-nuclear assets.

The time to do so is now. Nuclear war simulations—even open-source ones—are now widely available and can be a powerful tool.

The United States should use them.

Leo A. Keay is a PhD Candidate in the Department of Defence Studies at King’s College London. **Robert Peters** is Senior Research Fellow for Strategic Deterrence in the Douglas and Sarah Allison Center for National Security at The Heritage Foundation.

Endnotes

1. Leo A. Keay and Robert Peters, "Limited Nuclear War Over Taiwan: An Initial Exercise," Heritage Foundation *Issue Brief* No. 5402, forthcoming.
2. See Glenn A. Kent and David E. Thaler, *First-Strike Stability: A Methodology for Evaluating Strategic Forces* (Santa Monica, CA: RAND Corporation, 1989), and Austin Long, "U.S. Nuclear Strategy Toward China: Damage Limitation and Extended Deterrence," in Caroline Dorminey and Eric Gomez, eds., *America's Nuclear Crossroads: A Forward-Looking Anthology* (Washington, DC: Cato Institute, 2019), pp. 47–56.
3. For the baseline dataset, see "Azure Dragon II Results–Baseline," https://docs.google.com/document/d/1pVKRbfog80IG1mJTIYhcQzF_RHmTHbiSkk8-mIFgqhl/edit?usp=sharing (accessed December 19, 2025).
4. For the World A dataset, see "Azure Dragon II Results World A," <https://docs.google.com/document/d/1gYUNX344CreENypeA-outos8GuLSfDW59vpDLmtY00M/edit?usp=sharing> (accessed December 19, 2025).
5. For the World B dataset, see "Azure Dragon II Results World B," https://docs.google.com/document/d/12LSyGqgfmynzndvhYcLAzd7IHHq8a_yHjyLQud6TOwtQ/edit?usp=sharing (accessed December 19, 2025).
6. For the World C dataset, see "Azure Dragon II Results World C," <https://docs.google.com/document/d/197ER5ARQeuXG6Ig7IFUCSNnoQryuZxousOpPBouB1vc/edit?usp=sharing> (accessed December 19, 2025).
7. Brad Roberts, "On Theories of Victory, Red and Blue," Center for Global Security Research *Livermore Papers on Global Security* No. 7, June 2020, <https://cgsr.llnl.gov/sites/cgsr/files/2024-08/CGSR-LivermorePaper7.pdf> (accessed December 19, 2025), and Vipin Narang and Pranay Vaddi, "How to Survive the New Nuclear Age," *Foreign Affairs* (July/August 2025), <https://www.foreignaffairs.com/united-states/how-survive-new-nuclear-age-narang-vaddi> (accessed December 19, 2025).
8. J. Michael Legge, "Theater Nuclear Weapons and the NATO Strategy of Flexible Response," Report No. R-2964-FF, RAND Corporation, 1983, <https://www.rand.org/pubs/reports/R2964.html> (accessed December 19, 2025).