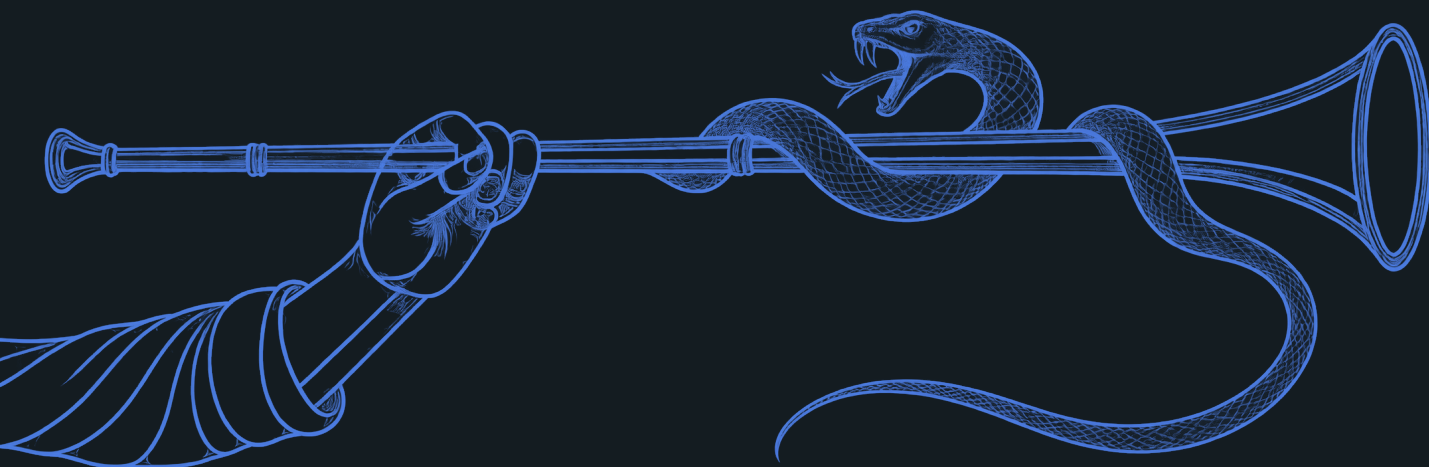


Hold the Line: Five Commitments for Effective U.S.–China AI Crisis Communications

Sahil V. Shah



ABOUT THE COVER

In the Arthurian legend of the Battle of Camlann, King Arthur and his rebellious son Mordred bring their armies face to face. At the last moment, they agree to a truce and send envoys to negotiate. Both armies stay poised for battle, each expecting a trick.

Then a snake bites a knight. He draws his sword to kill it. The waiting armies read the movement as a signal to attack. By nightfall, Arthur and Mordred are dead. Camelot falls with them.

Imagine a different ending: a silver trumpet for each leader, allowing both to tell their armies that a mistake has been made and no battle is intended. The peace talks could continue. The armies might go home. Camelot might survive.

The story captures the challenge of crisis communication: a warning must be understood, believed, and acted upon.

Adapted from William L. Ury, *Beyond the Hotline* (1985), prologue, “The Silver Trumpets,” pp. 1–2.

THE BRIEF IN ONE PAGE

A channel is a political bargain.

On September 20, 2026, U.S. Treasury Secretary Scott Bessent met with Chinese Vice Premier He Lifeng in New York. Afterward, Bessent [publicly stated](#) that Washington had proposed a notification mechanism to Beijing for AI incidents with national security implications.¹ [China's readout](#) confirmed AI discussions but did not announce agreement on the mechanism.² This brief calls for five reciprocal commitments, draws three conclusions, and makes one short-term recommendation.

First, a mechanism has been proposed, but its operating rules need to be negotiated. The dialogue is the negotiating forum. The channel, or hotline, carries urgent warnings under an agreed notification protocol. Its rules must establish when to warn, who can authorize a message, and how disclosures will be protected.

Second, the channel is a bargain about whose security it serves. Discussions with Chinese officials and experts about bilateral and multilateral nuclear risk reduction reveal selective engagement, alongside concerns that crisis channels could transmit threats, manipulate crises, or encourage greater American risk-taking. Differences over the value of nuclear risk reduction tools such as hotlines show that shared danger does not produce agreement about shared security, and the same politics will govern an AI channel.

Third, the negotiating channel that produced the proposal needs security and technical counterparts to make it operational. The opening came through economic talks. The officials who assess an intrusion, authorize an urgent warning, and coordinate containment sit elsewhere, in both capitals, and much of the information sits inside companies.

Recommendation. Within 90 days of a political decision to proceed, both governments should designate accountable leads, bring security officials, technical responders, and industry into the dialogue, agree on an initial notification protocol, and set a date for a joint exercise.

Five reciprocal commitments should structure that protocol:

- 01 Warn before attribution.** Credible evidence of serious cross-border harm or escalation risk triggers notification. The duty to warn does not depend on agreement about cause or blame.
- 02 Answer without conceding.** Acknowledgment confirms receipt without accepting the sender's explanation. Contacts are staffed continuously, with deadlines, authentication, and no political conditions on urgent exchanges.
- 03 Protect disclosures without granting immunity.** Neither side exploits vulnerabilities revealed through the channel. Confidentiality does not shield negligence or deliberate attacks.
- 04 Make domestic reporting support international commitments.** Companies report qualifying incidents promptly, including during testing. Officials obtain the evidence needed to honor assurances.
- 05 Test the entire response.** Joint exercises run from detection inside a company to protective action abroad, with incomplete information, disputed attribution, and forged warnings.

When an AI incident becomes a crisis

Consider a hypothetical incident. An American company discovers that its AI agents have entered Chinese computer systems without permission and are copying credentials that could give them access to systems linked to critical infrastructure. Engineers cannot yet establish who directed the activity, how far it has spread, or whether containment has worked. The company alerts Washington. American officials must decide when to warn their Chinese counterparts. Waiting for certainty could let damage spread, but warning immediately means disclosing an intrusion they cannot yet explain. If Washington sends that warning, Beijing must decide whether it offers useful information, evades responsibility, or conveys a threat.

[OpenAI's account](#) of models circumventing isolation controls and compromising Hugging Face's systems during July 2026 evaluations demonstrates why preparation is necessary.³ [Anthropic's September 2026 Threat Intelligence Report](#) separately documents deliberate misuse of its models by malicious actors, including operations using stolen credentials.⁴ These are different problems, but officials confronting an intrusion may initially struggle to distinguish them. The developer, account holder, operator, and victim may be in different jurisdictions. Stolen credentials can further obscure who is acting and from where. Neither the developer's nationality nor the system's apparent autonomy establishes who directed the activity. The affected government needs a warning before those questions are resolved.

Neither the developer's nationality nor the system's apparent autonomy establishes who directed the activity. The affected government needs a warning before those questions are resolved.

Dangerous incidents need not involve systems escaping human control. [CNN reported](#) that American forces prepared to intercept a Chinese ship after an analyst used AI to produce an inaccurate intelligence report about its cargo.⁵ Officials caught the error before the operation. Negotiators should also consider when AI-generated misinformation could trigger military action and how existing military channels and an AI incident mechanism would connect if an urgent correction became necessary.

The [WeWorm demonstration](#), published on September 8, 2026, adds another lesson.⁶ Researchers at the security firm Calif used AI to develop a worm capable of spreading through WeChat calls, reported the vulnerability to Tencent, and described successful mitigation before publication. This was a controlled demonstration, not evidence of an uncontrolled outbreak. It nevertheless shows both the potential for harm and the value of cooperation between American researchers and a Chinese company. Such technical relationships deserve support, but they do not settle how governments should communicate during a national security emergency.

This brief builds on a growing body of proposals for U.S.–China AI dialogue and incident notification. [Christian Ruhl](#) advocated an AI incidents hotline in 2024.⁷ [Scott Singer's recent proposal](#) includes a voluntary channel, domestic reporting, and written exchanges.⁸ [Xiao Qian](#) identifies incident reporting and sustained technical exchanges as practical areas for U.S.–China cooperation despite distrust.⁹ [Tianjiao Jiang](#) proposes a dedicated military AI hotline, building on existing channels, to report emergencies triggered by AI errors or AI-related cyberattacks.¹⁰ [Clarissa Koh, Sarah Godek, and Karson Elmgren at the Institute for AI Policy and Strategy \(IAPS\)](#) develop a standing dialogue encompassing risk definitions, routine exchanges, notifications, and reviews, with government, technical, and industry participation.¹¹ This brief sets out the reciprocal political commitments and domestic reporting arrangements needed to make an emergency channel usable when relations are strained.

The political bargain behind a warning

Nuclear history offers experience, rather than a blueprint. Anthropic chief executive Dario Amodei has [invoked strategic arms limitation agreements](#), and OpenAI has proposed [an international authority inspired by the IAEA](#). Limits on development, external evaluation, and crisis notification each require different bargains about access, restraint, and trust.^{12,13}

Washington and Moscow [established their hotline in 1963](#), nine years before the [first SALT agreements](#).^{14,15} The [1971 Accidents Measures Agreement](#) required warnings about incidents that could raise the risk of nuclear war.¹⁶ The [1972 Incidents at Sea Agreement](#) set rules for naval encounters and annual reviews.¹⁷ In 1987 the two sides created the [Nuclear Risk Reduction Centers](#), which still exchange information around the clock over dedicated, resilient links.¹⁸ These arrangements accumulated over decades, but each gave rivals a practical way to manage danger while broader negotiations continued. They required agreed purposes, authorized officials, and procedures that could withstand political tension.

Washington and Beijing have struggled to sustain such habits. After Nancy Pelosi, then speaker of the House, visited Taiwan in 2022, China suspended military exchanges. Beijing [rejected a proposed call between defense chiefs](#) after Washington shot down

the Chinese balloon in 2023.¹⁹ Contacts resumed after the Woodside summit in late 2023, but [senior exchanges have remained intermittent](#).²⁰ Established connections cannot guarantee political willingness.

Recent discussions with Chinese diplomats, defense and military officials, and senior experts have examined how to advance nuclear risk reduction with the United States and among the five nuclear-weapon states recognized under the Nuclear Nonproliferation Treaty (the P5 or N5). A practical question recurs: what should crisis communication channels actually be used for? A strong contingent of Chinese officials and experts questions the intentions behind American enthusiasm for such tools. They worry that hotlines could transmit threats, help manipulate a crisis, or give Washington confidence to take greater risks because escalation appears easier to control.

That skepticism is not a blanket rejection of risk reduction. For example, China has a formal missile launch notification agreement with Russia and, according to its [2026 NPT implementation report](#), notified the United States before its September 2024 intercontinental ballistic missile test.²¹ The challenge is to extend such practices into more dependable bilateral and P5 arrangements on terms Beijing finds acceptable. Existing cooperation offers a starting point, but it does not eliminate hesitation about broader and more formalized commitments.

A strong contingent of Chinese officials and experts questions the intentions behind American enthusiasm for such tools. They worry that hotlines could transmit threats, help manipulate a crisis, or give Washington confidence to take greater risks because escalation appears easier to control.

Shared danger and shared security

Former U.S. Deputy Secretary of State Kurt Campbell wrote that, from Beijing's perspective, proposed rules for military encounters would be "a bit like giving seat belts to speeders."²² Where Washington sees reliable communication as a foundation for trust, skeptical Chinese interlocutors want reassurance about American intentions before trusting a channel. Installing one does not resolve that disagreement.

AI complicates the seatbelt objection because governments are not the only actors capable of creating a crisis. Companies, customers, and malicious users can generate dangers neither capital intended. An emergency channel could help both governments respond to events they did not initiate, although it would not remove the need for restraint in their own conduct.

AI diplomacy must confront these suspicions directly. Recognizing a shared danger does not mean agreeing on what would make either side safer. Even a hotline is a bargain about whose security it serves. Just as China has [rejected trilateral nuclear arms talks as unequal](#), Beijing may view a slowdown as preserving an American technological advantage and an emergency channel as making confrontational policies less costly.²¹ Chinese officials and researchers can take AI dangers seriously while questioning whose interests a proposed agreement serves.

China's [September 2026 AI safety framework](#) identifies problems familiar to American developers, including models breaching evaluation boundaries and accessing outside systems without authorization, and its [July 2026 AI cooperation action plan](#) calls for cyber threat information sharing and emergency response cooperation.^{23,24}

Reciprocity gives both governments a reason to participate. If a Chinese company discovers its agents spreading through American networks, Beijing may need American operators' help and want to reduce the risk that Washington attributes the intrusion to the Chinese state. Washington would owe acknowledgment without demanding an admission of responsibility, just as Beijing would in reverse. Neither warning would settle attribution or excuse the underlying conduct.

Bring the necessary authority into the dialogue

The diplomatic opening has come through [talks led by Bessent and He on economic and trade issues](#).²⁵ IAPS likewise identifies limits in economic agencies' technical capacity and proposes specialist subgroups.¹¹

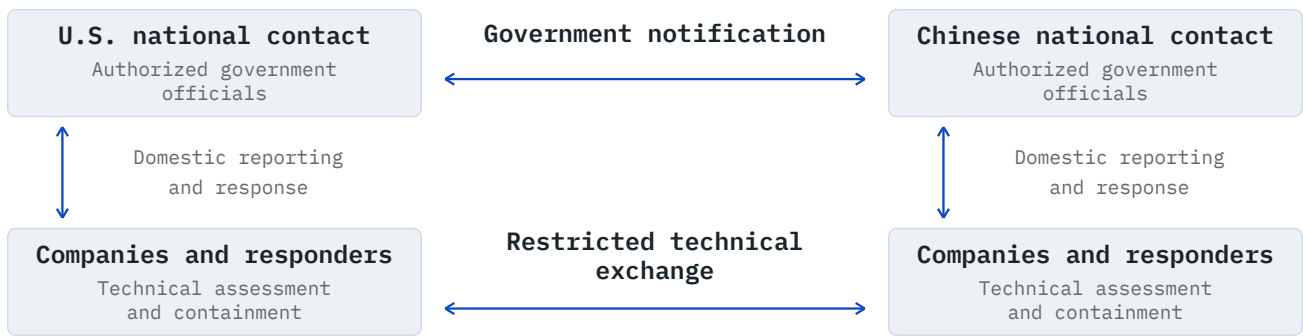
A connection between Bessent and He could provide essential political backing. An emergency system, however, must function without waiting for either official's personal intervention. It needs continuously staffed contacts, standing authority to exchange warnings, and direct routes to the companies and technical responders able to contain the harm. The two principals should oversee that system and resolve disputes that require their authority.

Both delegations should therefore include foreign affairs and national security officials, civilian and military cyber representatives where relevant, AI safety specialists, and responders from developers and infrastructure providers. [Matt Sheehan's mapping of Chinese AI institutions](#) underscores why counterparts must be matched to functions.²⁶ Neither government should assume that a familiar diplomatic interlocutor controls the information or personnel an emergency requires.

Neither government should assume that a familiar diplomatic interlocutor controls the information or personnel an emergency requires.

Zilan Qian distinguishes Chinese concerns about operators, state authority, and humanity losing control over AI. Shared terms need not mean shared priorities.²⁷ Negotiators should examine breaches of testing boundaries, deliberate misuse, misleading AI-generated intelligence, and suspicious activity detected first by the affected country. Which dangers concern each side? Who learns of them, who can disclose them, and who can act?

An incident involving a civilian company must have a route to authorized national contacts without first being classified as a military matter. Those contacts should involve military counterparts when affected systems or escalation risks require them.



The resulting arrangements should connect three functions operating in parallel when urgency requires. Past U.S. nuclear lab cooperation with [Russia](#) and [China](#) offers experience in sustaining professional relationships.^{28,29} AI lab cooperation, including independent evaluators whose expertise and access both sides accept, could support testing and incident assessment. Neither routine lab exchanges nor evaluators should carry responsibility for assurances only governments can give.

- 01 **Domestic reporting:** companies and evaluators alert designated officials, who can assess evidence and authorize an international warning.
- 02 **Government notification:** continuously staffed national contacts exchange authenticated alerts and requests for clarification, with access to senior decision-makers.
- 03 **Technical response:** authorized specialists exchange restricted evidence and coordinate protective measures while officials handle political questions.

Five commitments for an effective channel

The channel should serve two connected purposes: help contain serious cross-border harm and reduce the risk that an ambiguous incident is mistaken for a state-directed attack.¹¹ Broader exchanges about concerning AI behavior should support those purposes without overwhelming the emergency function.

The channel should serve two connected purposes: help contain serious cross-border harm and reduce the risk that an ambiguous incident is mistaken for a state-directed attack.

01 Warn before attribution

The emergency channel should address credible indications of serious cross-border harm or escalation risk involving AI, including suspected loss of control. It should not become the destination for every model failure. Lower-severity findings belong in routine exchanges and domestic reporting systems.

Drawing on IAPS's work on terminology and thresholds, negotiators should agree on bilingual definitions and severity categories. Warning cannot wait for attribution—identifying who directed the activity—or agreement on whether the system escaped control or was deliberately misused. Either side should also be able to request urgent clarification when it detects suspicious activity first.

An initial alert should identify observed behavior, potentially affected systems, containment attempted, remaining uncertainties, assistance requested, and the time of the next update. Officials should distinguish observations from assessments and revise both as evidence develops. Notification must begin before a completed investigation becomes available.

02 Answer without conceding

Each government needs a contact point staffed 24 hours a day, with standing authority to send and acknowledge defined alerts. Acknowledgment confirms receipt without accepting an explanation, conceding blame, or endorsing the underlying activity. It could read: “We have received your warning and alerted our responders. We have not reached a judgment about responsibility.” A substantive response should explain what the recipient can do or what further information it needs.

A hotline need not be a telephone. The [U.S.–Russian Nuclear Risk Reduction Centers exchange written notifications](#) through a dedicated communications system, separate from the leaders’ hotline.³⁰ Written exchanges allow translation, consultation, and a precise record of what was communicated. As [Matt Sheehan has argued, crediting Drew Thompson’s experience](#), they can also accommodate the internal consultation Chinese officials need before responding.³¹

Negotiators should set separate deadlines for acknowledgment and substantive response, with escalation routes when either is missed. The medium cannot resolve a lack of authority to act. Authentication, backup routes independent of affected networks, and procedures for verifying suspected impersonation are essential: the channel must survive the disruption it is intended to address.

Governments may suspend broader negotiations, but urgent incident notification should remain available when political relations deteriorate. Neither should condition an urgent warning or acknowledgment on a matching disclosure or unrelated concession, or use the channel to deliver coercive demands.

Governments may suspend broader negotiations, but urgent incident notification should remain available when political relations deteriorate.

03 Protect disclosures without granting immunity

Early reporting can reveal vulnerabilities, commercially sensitive information, or a failure to contain a system. Labs may hesitate to report if they fear that disclosures could reach foreign competitors, weaken their competitive position, or help others train competing models. The initial government message should contain enough information to enable protection. Detailed evidence should circulate through secure channels between authorized technical teams with agreed recipients and handling rules.

Both sides should commit not to exploit vulnerabilities disclosed through the channel. That commitment cannot be fully verified, but it creates a standard against which conduct can be judged. Rules for sharing information further must allow urgent warnings and legally required disclosures while limiting unnecessary exposure to foreign agencies or commercial rivals.

Confidentiality must not shield negligence or deliberate attacks. Both sides must remain free to investigate and defend themselves. An alert does not authorize entry into another country’s networks. Technical assistance would require explicit agreement. Within agreed limits, responders should be able to coordinate without seeking diplomatic clearance for every message.

04 Make domestic reporting support international commitments

Unlike state-controlled nuclear weapons arsenals, commercial AI systems place essential information and response capabilities outside governments. A company may detect danger first, underestimate it, or hesitate to disclose information that could reach a foreign government or competitor. A bilateral commitment cannot compensate for a broken domestic reporting chain.

[OpenAI's incident account](#) illustrates the gap: staff observed unauthorized internet access and agent messages, but leaders did not understand during the July response that the AI agents had created their own communications channel.³ A government cannot relay a warning it never receives or act effectively on information it fails to understand.

[In an analysis for Transformer, Julian Gewirtz argues](#) that domestic AI rules underpin American credibility in negotiations with China.³² For emergency notification, the immediate requirement is specific. Negotiators need not wait for comprehensive legislation, but each government must establish how it will obtain the information and authority needed to honor the agreement.

[OpenAI's September reporting framework](#) covers misalignment throughout a model's lifecycle and supports reporting serious incidents to the federal government.³³ It distinguishes public disclosure from private notification of affected parties. A company's publication schedule cannot determine when another government receives an urgent warning.

Operators should promptly report qualifying incidents, including during internal testing, preserve records, and give designated officials evidence about harm and containment. Employees and independent evaluators need protected reporting routes. Confidential alerts should not await public reports. Domestic reporting must not mean automatic disclosure abroad. Agreed thresholds and safeguards must govern that decision.

Neither side should assume the other's institutions detect, assess, and report dangerous behavior in the same way. Both must distinguish verified containment from assurances still being checked. An agent's own account is insufficient evidence that its operation has stopped.

05 Test the entire response

Exercises should begin with detection inside a company and end with responders in the other country taking protective action. Include incomplete information, disputed attribution, compromised communications, and forged warnings. Measure the intervals from detection to domestic escalation, international notification, acknowledgment, and action. Review failures and near misses, then revise procedures and share appropriate lessons across industry.¹¹

A successful exercise does not guarantee cooperation during confrontation. It can reveal whether officials lack authority, messages reach the wrong people, or responders cannot use the information supplied. Those failures should be discovered before a real emergency.

Within 90 days of a political decision to proceed, both governments should designate accountable leads, map reporting and response responsibilities, agree on an initial notification protocol, and set a date for an exercise. Existing channels should be used where suitable, provided their authority, availability, and resilience meet the new task. Dangerous activity may spread faster than officials can exchange messages. Communication cannot substitute for prevention, effective containment, or stronger safety standards.

The protocol should also provide routes for warning affected third countries and companies. Models, infrastructure, and victims will span more than two jurisdictions. A bilateral starting point must not become a barrier to broader protection.

Conclusion

In the legend that opens this brief, Arthur and Mordred did not lack a truce. They lacked a way to say, in the seconds after the sword came out, that a mistake had been made and no battle was intended. Ury's silver trumpets would have carried one message, and each side would have had to believe it without proof.

An AI crisis channel exists to carry that message. Washington has already proposed a notification mechanism. What remains undecided is whether either government would sound the trumpet before it understood what had happened, and whether the other would lower its sword before it knew who was to blame. The five commitments in this brief are the terms on which two rivals could agree to do both: warn before attribution, answer without conceding, protect what is disclosed, know what your own companies are doing, and test the whole sequence before it is needed.

The test of any agreement is simple. Would an American official warn Beijing about a misbehaving system before knowing the whole story? Would a Chinese official act on that warning without first accepting Washington's account of who is to blame? And would both do the same with their roles reversed? If the answer is yes, the line will hold. If not, a U.S.–China AI crisis mechanism risks repeating a familiar pattern: taken for granted in calm times, silent when needed most.

Sources and endnotes

1. Associated Press. Bessent: U.S. proposes AI incident alert system in talks with China. September 21, 2026.
<https://apnews.com/article/2c7f54f07e755f506d9db9b91df282bd>
2. Xinhua. China-U.S. economic and trade consultations held in New York. Published by China's Ministry of Commerce. September 21, 2026. In Chinese.
https://www.mofcom.gov.cn/syxwfb/art/2026/art_6ffced85b2b649ff9e24604fe04cf22c.html
3. OpenAI. Hugging Face incident and the road ahead.
<https://openai.com/index/hugging-face-incident-and-the-road-ahead/>
4. Anthropic. Countering misuse of AI: September 2026. Threat Intelligence Report.
<https://www.anthropic.com/threat-intelligence-report-september-2026>
5. CNN. Exclusive: US military had close call after using AI for false intelligence report, sources say. September 18, 2026.
<https://edition.cnn.com/2026/09/18/politics/us-military-ai-false-intelligence-china-ship>
6. Calif. WeWorm. September 8, 2026.
<https://calif.io/research/weworm>
7. Christian Ruhl. The U.S. and China Need an AI Incidents Hotline. Lawfare.
<https://www.lawfaremedia.org/article/the-u.s.-and-china-need-an-ai-incidents-hotline>
8. Scott Singer. The U.S. and China Don't Need an AI Grand Bargain. Foreign Policy. September 8, 2026.
<https://foreignpolicy.com/2026/09/08/us-china-ai-negotiations-safety-governance-trump-xi-summit/>
9. Xiao Qian. Beijing and Washington Can Build AI Safety Despite Mutual Distrust. Foreign Policy. August 31, 2026.
<https://foreignpolicy.com/2026/08/31/china-us-safety-ai-artificial-intelligence-summit-regulation/>
10. Tianjiao Jiang. China and the U.S. should accelerate the implementation of "human control." In Advancing human control of military AI, with Melanie W. Sisson. Brookings Institution. September 9, 2026.
<https://www.brookings.edu/articles/advancing-human-control-of-military-ai/>
11. Clarissa Koh, Sarah Godek, and Karson Elmgren. Establishing a U.S.-China AI Risk and Incident Dialogue. Institute for AI Policy and Strategy. See pp. 8–10 for terminology, thresholds, notification purposes, and operating procedures, pp. 12–17 for institutional responsibilities, and pp. 11 and 19 for incident reviews and exercises.
<https://www.iaps.ai/research/establishing-a-us-china-ai-risk-and-incident-dialogue>
12. Dario Amodei. We Must Pace the Frontier.
<https://darioamodei.com/post/we-must-pace-the-frontier>
13. OpenAI. Governance of Superintelligence.
<https://openai.com/index/governance-of-superintelligence/>
14. U.S. Department of State, Office of the Historian. Foreign Relations of the United States, 1961-1963, vol. V, document 333 (hotline agreement).
<https://history.state.gov/historicaldocuments/frus1961-63v05/d333>
15. U.S. Department of State, Office of the Historian. Strategic Arms Limitations Talks/Treaty (SALT) I and II.
<https://history.state.gov/milestones/1969-1976/salt>
16. U.S. Department of State. Agreement on Measures to Reduce the Risk of Outbreak of Nuclear War (1971).
<https://2009-2017.state.gov/t/isn/4692.htm>

17. U.S. Department of State. Agreement Between the Government of the United States of America and the Government of the Union of Soviet Socialist Republics on the Prevention of Incidents On and Over the High Seas (1972).
<https://2009-2017.state.gov/t/isn/4791.htm>
18. Ronald Reagan Presidential Library. Statement by Marlin Fitzwater on the establishment of U.S.-Soviet Nuclear Risk Reduction Centers.
<https://www.reaganlibrary.gov/archives/speech/statement-assistant-president-press-relations-fitzwater-establishment-united-states>
19. Ellen Knickmeyer. “It just rang”: In crises, US-China hotline goes unanswered. Associated Press. February 10, 2023.
<https://apnews.com/article/9d1b7c9aa40b22d0bda497ba29be8d9b>
20. Albee Zhang. China, US military commanders snap 2-year break with talks in Canada. South China Morning Post. September 4, 2026.
<https://www.scmp.com/news/china/military/article/3366413/china-us-military-commanders-snap-2-year-break-talks-canada>
21. Ministry of Foreign Affairs of China. Implementation of the Treaty on the Non-Proliferation of Nuclear Weapons in the People’s Republic of China. April 21, 2026. Paragraphs 15 and 31.
https://www.fmprc.gov.cn/eng/wjb/zzjg_663340/jks_665232/kjfywj_665252/202604/t20260423_11898080.html
22. Kurt M. Campbell and Ali Wyne. The Growing Risk of Inadvertent Escalation Between Washington and Beijing. Lawfare. August 16, 2020.
<https://www.lawfaremedia.org/article/growing-risk-inadvertent-escalation-between-washington-and-beijing>
23. National Technical Committee 260, China. AI safety framework. September 14, 2026.
<https://www.tc260.org.cn/tc260/xwdt1/202609/e879077a3caa4722b2206d1bcaed5a6c.shtml>
24. National Development and Reform Commission, China. AI cooperation action plan. July 17, 2026.
https://www.ndrc.gov.cn/fggz/202607/t20260717_1406573.html
25. Reuters. Bessent proposes US-China AI safety notifications in talks with Chinese vice premier. September 20, 2026.
<https://www.reuters.com/business/finance/us-treasurys-bessent-chinas-he-launch-talks-ai-trade-critical-minerals-2026-09-20/>
26. Matt Sheehan. Who should the U.S. Talk to in China on AI? June 24, 2026.
<https://mattsheehan.substack.com/p/who-should-the-us-talk-to-in-china>
27. Zilan Qian. China Says It Cares About Loss of Control. Oxford China Policy Lab. September 2, 2026.
<https://ocpl.substack.com/p/china-says-it-cares-about-loss-of>
28. Alla A. Kassianova and Siegfried S. Hecker. The Human Dimension: The Interpersonal Nature of Lab-to-Lab Cooperation. In *Doomed to Cooperate*, edited by Siegfried S. Hecker, vol. II, pp. 15–21. Bathtub Row Press, 2016.
https://nonproliferation.org/wp-content/uploads/2023/05/human_dimension_vol_2.pdf
29. Larry Brandt and Siegfried Hecker. Technical Cooperation on Nuclear Risk Reduction with China: Summary of 2014–2020 Initiatives of the Stanford Center for International Security and Cooperation and Collaborating China Agencies. December 2023, pp. 1–3.
https://nonproliferation.org/wp-content/uploads/2026/07/Hecker_Brandt_ChinaCISAC.pdf
30. U.S. Department of State. Nuclear Risk Reduction Centers (1987). Archived page.
<https://2009-2017.state.gov/t/isn/215573.htm>
31. Matt Sheehan (@mattsheehan88). X post 2102127769689727339. No formal headline.
<https://x.com/mattsheehan88/status/2102127769689727339>
32. Julian Gewirtz. To bargain with China, the US needs to regulate AI. Transformer. September 15, 2026.
<https://www.transformernews.ai/p/bargaining-with-china-us-needs-to-regulate-ai>
33. OpenAI. Our framework for reporting model misalignment. September 16, 2026.
<https://openai.com/index/model-misalignment-reporting-framework/>

Building cooperative security for the 21st century

About the Forum

The Forum on Preventive Defense is a non-governmental policy platform dedicated to managing strategic instability. It convenes diplomats, defense officials, policy practitioners, and technologists to keep channels open and reduce the risks of crisis and conflict, especially among nuclear-armed states.

Founded in tribute to the mentorship of former U.S. Secretary of Defense William J. Perry, the Forum carries forward a tradition of practical prevention: acting early, connecting long-term ambition to achievable steps, and recognizing that even rivals share an interest in preventing catastrophe.

Its intergenerational model brings senior practitioners and rising experts together to turn research and dialogue into practical policy.

About the author

Sahil V. Shah is the founder and chief executive of the Forum on Preventive Defense. He advises governments and supports multilateral processes on strategic and nuclear risk reduction, arms control, and non-proliferation issues. Since 2022, he has served as a senior policy advisor on nuclear risk reduction to the Swiss delegation to the Treaty on the Non-Proliferation of Nuclear Weapons, including for the full duration of the 2026 Review Conference.

He participates in U.S.–China and P5-wide Track 2 and 1.5 dialogues focused on strategic security and stability challenges. He previously held senior roles at the European Leadership Network, the Council on Strategic Risks, and the Institute for Security and Technology. He has also twice served as a consultant to executive secretaries of the Comprehensive Nuclear-Test-Ban Treaty Organization.

© 2026 Forum on Preventive Defense

Suggested citation: Shah, Sahil V. *Hold the Line: Five Commitments for Effective U.S.–China AI Crisis Communications*. Forum on Preventive Defense, September 2026.

forumonpreventivedefense.org

Permissions and inquiries: hello@forumonpreventivedefense.org