SECURITY INDEX

№ 7 (12) | 2020

Strategic (In)security: Perspectives from the United States

In this issue

FROM EDITOR

PIR Center continues to publish policy papers, which were prepared for a joint seminar on reducing nuclear risks during great power competition, which it co-organized together with the Center for Strategic and International Studies (CSIS). As it appears that such competition is already underway, we decided to release two policy memos originally prepared for the seminar under one cover "*Strategic (In)Stability: Perspectives from the U.S.*". As discussed by the authors, there is some overlooked potential for constructive engagement between Russian and the United States with regards to arms control and emerging technologies.

The first essay is authored by Gen. Peter Zwack, who previously served as the U.S. Defense Attache to the Russian Federation. Although he poses as a seasoned "Cold Warrior", his recommendations are soberly realistic, with it being possible for a Russian expert to put a signature to most of his points. Gen. Zwack is particularly vocal in calling for military-to-military contacts in order to reduce the potential for misunderstanding each other`s motives and rationales. As the one who witnessed the late days of the deadly confrontation between the Soviet Union and the United States, he recognizes the perils of emerging technology which "can lead us further downwards".

Dealing with the nexus between challenges of modern technologies to strategic stability is the second piece penned by Dr. Margaret Kosal from Georgia Tech. As it was pointed out during the seminar in December, the situation in this domain (as well as in outer space) resembles the early days of the original Cold War, when the technology was being explored and applied militarily with little or no understanding how to tame it with political means. Dr. Kosal does not fall into the temptation of rushing to conclusions and recommendations, and her call to build on the experience of the traditional arms controls without confining ourselves to its limits resonates with our thinking.

As it has been the case previously, pandemics come and go, but the factors shaping the strategic environment stay, and we should never stop thinking of how to prevent the strategic stability from becoming instability. And our hope is that this occasional paper will contribute to this discussion.

Vladimir A. Orlov

Key findings

To contents

Peter Zwack. Reducing Nuclear Risk During Great Power Competition

- **First, any solution must start with the joint leadership of the Russia Federation and the United States.** The world is watching what we do and how we work together or against one another. It's our joint responsibility to set the tone and supply a model of cooperation – no one else will.
- **With the treaties already renounced disappeared critical US-Russian personal "contact points," where near daily eye-to-eye discussion occurred among numerous diplomats, scientists, engineers and military staff from both countries.** Those multi-level dialogues demystified and dedemonized one another, built trust even with disagreement and led to major breakthroughs.
- **Presidential statements alone in this difficult, distracting political climate will not do it.** The US Secretary of Defense should meet with the Russian Minister of Defense t o help frame these critical issues from a national defense and security perspective. It has been a long time since they have specifically met in sustained dialogue.

Read the paper on PIR Center website

Margaret Kosal. AI & Global Security Environment

- **In the post-Cold War environment, the most technologically advanced military power no longer guarantees national security.** As nations and the international community look to the future – whether dominated by extremist groups co-opting advanced weapons in the world of globalized non-state actors or states engaged in persistent regional conflicts in areas of strategic interest – new adversaries and new science and technology will emerge.
- **Currently all US operational systems require "human in the loop"** and are restricted in scope and nature, i.e., fixed anti-missile capabilities on ships, rather than general lethality.

As systems are developed and deployed wi th higher levels of autonomy, broader scope, and the ability to move independently, the calculus will change.

- **If human biases can impact machine learning outcomes for designing inorganic reactions, it's something to be cognizant of** for other – potentially more consequential – decision-making assisted by AI.
- To be clear, there is much to learn from and leverage in existing arms control and nonproliferation institutions. These starting points and history are valuable; they are not necessarily predictive, however. **Regardless, the challenges in this arena are primarily political rather than technical.**

Read the paper on PIR Center website

Acknowledgements

To contents

These papers have been produced for the joint PIR Center – CSIS project "Reducing nuclear risks during Great Powers Competition?" We thank our partners in CSIS for their cooperation and support for this publication.

About the authors

To contents

Peter Zwack - Brigadier General (Ret), writes as a Wilson Center Global Fellow at the Kennan Institute. He had a long 34-year US Army military career, culminating as the US Senior Defense Attache to the Russian Federation during the challenging period of 2012-2014.

Margaret Kosal is an associate professor at the Sam Nunn School of International Affairs and adjunct scholar to the Modern War Institute at the US Military Academy/West Point. Before joining the

Sam Nunn School of International Affairs, she was Science and Technology Advisor within the Office of the Secretary of Defense (OSD).

Kosal also served as the first liaison to the Biological and Chemical Defense Directorate at the Defense Threat Reduction Agency (DTRA). Kosal led the U.S. involvement in the NATO Nanotechnology for Defense Working Group.

Security Index Occasional Paper Series presents reports, analytical articles, comments and interviews that reflect the positions of Russian and foreign experts on the current challenges of global security and Russian policy in this sphere. The newsletter aims at providing clear analysis of global security problems and suggesting practical solutions.

Security Index Occasional Paper Series continues the *Security Index* journal published by PIR Center in 1994 – 2016. Authors and editors will be glad to receive comments, questions and suggestions on our e-mail address inform@pircenter.org.

Editor-in-Chief: *Vladimir Orlov*
Peer reviewer: Vadim Kozyulin
Managing editor: *Sergey Semenov*
Occasional Paper #7 (12)

Our e-mail address: **inform@pircenter.org**

Reducing Nuclear Risk During Great Power Competition

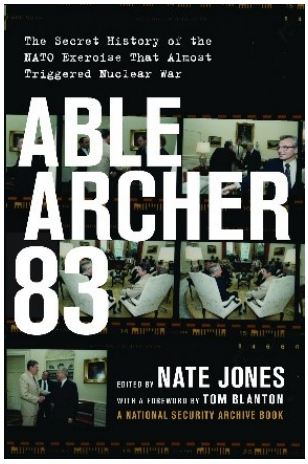
Peter Zwack

Over the past decade, overall US and Russian strategic stability and associated arms control measures have atrophied to dangerously low levels. The withering away of important arms reduction, verification and confidence-building initiatives—based mostly on "trust but verify" measures — has set relations among the United States, Russian Federation and several secondary nuclear states on a dangerous downward trajectory. If the trend continues, our already vulnerable world teeters even more on the precipice of a nuclear disaster.

Complicating an already dangerous situation are new and urgent concerns. Cyber and other technologies are being developed, which cannot be purely separated from nuclear weapons and their C2. New, difficult to detect and counter long-range precision weapons are being developed. Meanwhile disruptive conflict in the gray zone of today's 24/7 information space exponentially increases.

And, of course, there are the human and political components of the equation: the growing power of nuclear states that lack structured control measures; these include large nations (China, India, Pakistan), undeclared Israel, and emerging newcomers and aspirants (North Korea and Iran). How does one bring these countries into the arms control fold - especially when they are witnessing older nuclear powers backing away from existing arms control agreements?

Four Decades Observing the Landscape



As a military intelligence officer and later as a Soviet, then Eurasian Foreign Area I saw the creation of the On-Site Inspection Agency (OSIA) following the break-up of the US SR, and t he program's evolution into the Defense Threat Reduction Agency (DTRA) that still exists today.

Most treaties since then have fallen by the wayside even before today's slow-motion crisis. But it is today's rapid deterioration of remaining agreements that alarms me most. The crown jewel of strategic stability today—the New START Treaty signed in 2011is all that is left. Without update, or revision, it expires in February, 2021. If the New START disappears, without renewal, revision or replacement we are in nuclear freefall, breathtakingly vulnerable to an accident or incident that neither side wants.

Cold-bloodedly rational "Mutual Assured Destruction' (MAD) still very much lives today, but shakily so, with the fewer checks and balances of the past. To this author, serious progress in the strategic stability realm must include several interdependent tranches:

First, any solution must start with the joint leadership of the Russia Federation and the United States. The world is watching what we do and how we work together or against one another. It's our joint responsibility to set the tone and supply a model of cooperation – no one else will. More established nuclear powers such as France and the United Kingdom should follow suit – perhaps more wishfully along with China, India and Pakistan. Action must include renewed dialogue and negotiation on treaties and verification measures across a range of nuclear and nonnuclear technologies.

Younger generations must be educated. Hundreds of millions of today's global citizens were born after the end of the Cold War. Many take arms control for granted and assume that their leaders are maintaining a decades-old status quo that keeps the world safe. These younger people need to understand the dangers of allowing carefully crafted agreements to unravel with no modification or replacement in sight. Key to their education is showing them that the United States and Russia recognize the importance of a frank joint reappraisal of the current situation and are seen as committed to taking joint action.

We must recommit ourselves to constructive dialog, both formal and informal, to increase understanding and thereby avoid further demonizing one another. The roll call of canceled treaties and agreements linked to strategic stability is sobering to consider. Important as New START is, other programs we've lost are perhaps even more so when added up as a whole. Both conventional and nuclear-focused treaties, ranging from Intermediate-range Nuclear Forces (INF), Anti-Ballistic Missile (ABM), to Conventional Force Europe (CFE) and Cooperative Threat Reduction (CTR) are gone.

With those programs disappeared critical US-Russian personal "contact points," where near daily eye-to-eye discussion occurred among numerous diplomats, scientists, engineers and military staff from both countries. Those multi-level dialogues demystified and dedemonized one another, built trust even with disagreement and led to major breakthroughs.

Major remaining confidence building verification measures are at risk especially those linked to New Start, as well as the multi-national "Open Skies," Non-Proliferation Treaties (NPT) and the Comprehensive Nuclear Test Ban Treaty (CBCT).

We can't let new technology lead us down a dangerous path. The trust deficit between our nations grows wider. Without the ongoing contact that accompanies implementation of various treaties, agreements and associated dialogues, the potential for lethal misunderstanding greatly escalates. Crisis decision-making, always immensely challenging, must now be conducted within increasingly ambiguous messy and murky circumstances, thanks to the sheer mass and speed of information enabled by a continuously evolving cyber backbone that barely existed a generation ago. Add to the mix the uncharted territory of artificial intelligence (AI) and other "no-human-on the-joystick-or-button" technologies, and the potential for mistakes in dealing with a fast-breaking crisis between superpowers becomes absolutely frightening to consider.

Those of us of a certain age well remember several mistakes and incidents that could have led to catastrophe during the "analog era" Cold War. Disaster was averted only by the measured, rational thinking of key persons on both US and Soviet sides, including several operational level officers virtually at the nuclear tip-of-the-spear.

Today, the threat of a mechanical and/or functional glitch in either a weapons or detection sensor system is likely more dangerous than any premeditated national decision to initiate hostilities. Like dominos, a misperceived action or accident could unintentionally initiate a hair-trigger, cyber-fast tit-for-tat escalation that could rapidly engulf both sides and with them the entire planet.

A Few Recommendations

First, and most fundamentally, attack the stifling distrust that will cripple any viable future initiatives. If I were to choose one area to focus our mutual efforts on, it would be improving fundamental TRUST. Right now, extreme distrust exists at every level of our governments and societies. The network of scientists, bureaucrats and diplomats that built past treaties and agreements during the very distrustful Cold War era is but a shadow of its earlier self. Senior-level interaction exists between Moscow and Washington, but these periodic exchanges are not enough and barely scratch the surface. Thanks to lightning-fast cyberspace, a crisis in the Pacific, Arctic, Black Sea or Mediterranean—not just edgy Eastern Europe or Syria—could erupt locally and in the blink of an eye consume our national governments in Moscow or Washington before anyone has time or presence of mind to stop a lethal chain reaction.

Next, look at strategic stability through several different lenses.

- Are Long Range Precision Conventional weapons a threat to ICBMs? What to do about uncontrolled IRBMs?
- How can distrusting nations do credible cyber/AI arms control and regulation? Is it even possible? How to manage and mitigate cyber-fast crisis, especially the prospect of increasingly automated decision-making.
- Missile Defense (MD). Is this concept outdated? How does one protect allies such as NATO, Japan and South Korea against regional missile threats? Can updated MD be made relevant for new technologies including long-range precision weapons?
- How new technologies can paralyze or obfuscate C2 and decisionmaking?

The US-Russian governments and militaries must make this effort a priority. Presidential statements alone in this difficult, distracting political climate will not do it. Measures should

include:

- The US Secretary of Defense should meet with the Russian Minister of Defense to help frame these critical issues from a national defense and security perspective. It has been a long time since they have specifically met in sustained dialogue.
- With "eyes wide open" senior military of both nations—including at a minimum leaders and key staff members of the US Joint Chiefs, STRATCOM and NORTHCOM—should meet with Russian counterparts in the General Staff plus Strategic Aviation and Rocket Forces for frank, problem-solving discussions on these issues. Additionally, pragmatic leader-to-leader links should be reestablished between Russian and U.S. regional commands worldwide such as INDOPACOM and the Russian Eastern Military District, European Command with Western/Southern/Northern Fleet MDs (also with NORTHCOM) and Central Military Command with the Central Military District. Such contact—even in the face of official distrust and disagreement—would provide vital understanding of each other's activities and perspectives between leaders at these echelons, and military commands worldwide. Such personal relationships could be a critical first-phase breakwater in the event of a fast-breaking regional crisis, especially if accidental or incident based.
- Where possible, efforts must be made to bring China into some of these dialogues especially in the nuclear realm. US and Russian progress should not be dependent or held hostage to any Chinese lack of willingness or disagreement to participate.
- Exchange in 2020 of US bipartisan Congressional and Russian Duma delegations focused specifically on Strategic Stability.
- Finally expand mutual identification of transnational criminal cyber and other capabilities and activities that could confuse and corrupt strategic decision-making.

Russia and the US are still preeminent in nuclear issues and both understand better than anyone that a very dangerous new world is emerging. The US and Russia, both experienced nuclear practitioner nations, must find a way together to limit and mitigate these emerging threats or our nations and the entire world will be in freefall. We succeeded in doing this before, during the Cold War, and we need to do it again, with even more purpose and determination.

AI and Global Security Environment

Margaret Kosal

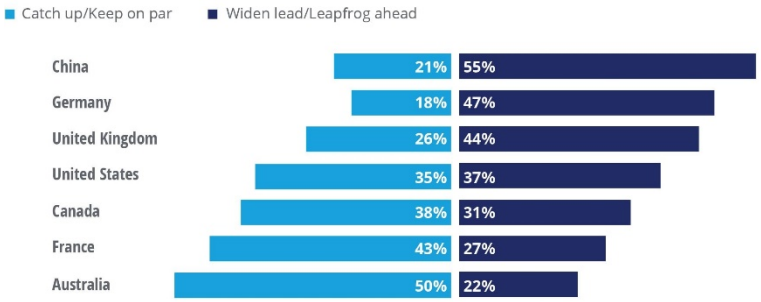
Disruptive technologies and emerging innovations within today's most cutting-edge science and technology (S&T) areas are cited as carrying the potential to revolutionize governmental structures, economies, and international security. Some have argued that such technologies will yield doomsday scenarios and that military applications of such technologies have even greater potential than nuclear weapons to radically change the balance of power.[1] While the suggestion that such emerging technologies will enable a new class of weapons that will alter the geopolitical landscape remains to be realized, a number of unresolved security puzzles have implications for international security, defense policy, governance, and arms control regimes. The extent to which these emerging technologies may exacerbate or mitigate the global security and governance challenges that states will pose in the future to global security interests will remain an integral question as policy-makers and leaders navigate the complex global environment

How, when, where, and in what form the shifting nature of technological progress may bring enhanced or entirely new capabilities, many of which are no longer the exclusive domain of a single nation-state, is contested and requires more cross-disciplinary thinking. Contemporary analyses of these emerging technologies often expose the tenuous links or disconnections among the scientific and technical realities and mainstream scholarship on national and international security.

In the post-Cold War environment, the most technologically advanced military power no longer guarantees national security. As nations and the international community look to the future – whether dominated by extremist groups co-opting advanced weapons in the world of globalized non-state actors or states engaged in persistent regional conflicts in areas of strategic interest – new adversaries and new science and technology will emerge. These new technologies and discoveries may significantly alter military capabilities and may generate new threats against military and civilian sectors. Greater strategic understanding of these game-changing technologies and the development of meaningful and testable metrics and models to help policymakers address the challenges of this complex global environment is needed.

FIGURE 5

AI early adopters in some countries are more likely to use AI to create a strong competitive advantage



Note: Not shown—respondents who take the more neutral stance that they're "edging slightly ahead."
Source: Deloitte State of AI in the Enterprise survey, 2nd Edition, 2018.

Deloitte Insights | deloitte.com/insights

Possible Challenges to Strategic Stability

The concept of strategic stability arose in the post-WWII nuclear policy realm in which military use of such weapons was recent memory. In the ensuing decades, it has become a cornerstone of national and international security and foreign policy by nuclear and non-nuclear states and cornerstone of deterrence.[2] Schelling and Wohlstetter-esque "stability of mutual deterrence" evokes strong connotations of stable and unstable equilibrium from the physical

sciences.[3] Strategic stability was all about surviving a first nuclear attack and then credibly being able to respond with a massive retaliatory nuclear strike and how that calculus critically affected geopolitics. The Cold War paradigm sought strategic stability through parity of nuclear arsenals in terms of capabilities, numbers, and conceptual permissiveness of limited nuclear war fighting and conformity of intent.

How, to what extent, and in what ways Artificial Intelligence (AI) may affect strategic stability is speculative. The concepts below are grounded in geopolitical and technical robustness, nonetheless they are intended to be illustrative rather than predictive.

Situational Awareness / ISR

As the limits of human capacity to process large streams of data, especially in time-sensitive environments, the risk of "data overload" increases. AI, particularly in the context of machine learning, is seen as valuable for data fusion from heterogeneous streams originating in large number of sensors, communications networks, and other electronic devices. Currently the US DoD Project Maven/Algorithmic Cross-Functional Team is a first attempt directed to identify and locate Daesh/ISIL fighters.

Command and Control (C2)/ Command Decision Support

Beyond situational awareness, another potential application of AI is to increase decision-making capacities. For example, the USAF Multi-domain Command and Control (MDC^2) system is meant to assign tasks to air, space, and cyber forces. The DARPA Artificial Intelligence Exploration (AIE) generates, tests, & refines hypotheses to assist human decision-making..

Cyber

AI has the potential to reduce uncertainty by helping make cyber networks more secure through detection of anomalies, identification of vulnerabilities, and potentially implement protective action (patch, isolate, self-heal, etc.) Examples include the DARPA 2016 Cyber Grand Challenge which reduced process to seconds from previous metric of days to detect cyber intrusions and the NSA's Sharkseer program, which monitors incoming email traffic to DoD servers for malware. Machine learning is also likely to be used for software verification and validation.

With respect to offensive cyber operations, AI may create vulnerabilities through introduction of incorrect training data as part of machine learning.

'Flash Crashes' or unexpected catastrophic failures are another concern with increasing incorporation of AI into complex, interconnected systems. Applying this to nuclear weapons and strategic stability can be done through the lens of "Normal Accidents" theory, originally proposed by Charles Perrow and applied to nuclear weapons by Scott Sagan.[4]

Autonomy

While much attention in popular press and at the international level has been given to autonomous systems, i.e., unmanned aerial vehicles, aka 'drones,' and lethality, the direction of the vector regarding increasing or decreasing stability is not resolved. **Currently all US operational systems require "human in the loop"** and are restricted in scope and nature, i.e., fixed anti-missile capabilities on ships, rather than general lethality. As systems are developed and deployed with higher levels of autonomy, broader scope, and the ability to move independently, the calculus will change.

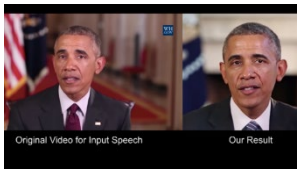
One area of particular concern is swarms, i.e., multiple independent autonomous systems that can synchronize and coordinate collective offensive and/or defensive maneuvers. Frequently these have been envisioned as large (n >10) formations of low-cost UAVs that might be used to overwhelm ground or ship-based defensive systems or troops. The technology to enable swarm tactics will require advances in AI for imagined scenarios to be realized.

Nuclear

The specific applications of AI to nuclear weapons directly often can take on a 'Dr.Strangelove"-esque motif. As far as implications for strategic stability, the application of AI that most often is mentioned is incorporation into launch on warning systems. This could result in a decreased decision-time by another nuclear state. Typical scenarios start with AI applied to machine vision and signal processing, which is then combined with autonomy and/or sensor fusion, to enable asymmetric capabilities for ISR, automatic target recognition (ATR), and technical guidance capabilities. Such capabilities could increase the likelihood that survivable forces (e.g., SLBMs and mobile missiles) could be targeted and even potentially destroyed, thereby also leading to increased plausibility of first strike.[5] It has been noted that such systems may undermine strategic stability even if state possessing such capabilities has no intention to use them,[6] as an adversary cannot be sure and may hedge.

Things to Watch Out For

Deep Fakes



Emerging video manipulation and fraudulent simulation technology that combines facial recognition with a neural network to allow its users to create fake monologues by public figures are referred to a 'deep fakes.' As this technology proliferates, it has the political to intensify political instability. By increasing the impact of misleading content, 'video spoofing' could lead to a rise in fake news, leadership imitation, and plausible deniability. As each of these scenarios can threaten key tenants of political stability—especially states with weak or

compromised political structures—it is within international security interests to prepare for and counter the threat posed by this emerging technology.

AI and related technology may need to be employed as countermeasures to authenticate video. Video watermarking techniques already exist that allow authors to embed signatures of authenticity in the video itself.[7] Additionally, steganalysis and statistical methods can be used to search video for digital irregularities that are indicators of post-creation modification;[8] improving and innovating such media forensic techniques is the objective of DARPA's current 'MediFor' project. Finally, checksum file verification programs can be used to ensure that a file in question is the same as the one it is supposed to be; platforms like ConceptCrypt take advantage of the immutability of the blockchain for this very purpose.

Hype

In geopolitics, rhetoric matters. It's not the only thing that matters nor often the most important, but it does matter. And therefore, one must be cognizant of hype. A prime example of this is the "Slaughterbots video," produced by an NGO seeking an international treaty to ban lethal autonomous systems.[9]

Will AI Replicate Human Biases, Stereotypes, & Prejudices?

As machine learning applications such as facial recognition are increasingly employed, study of how the training data may replicate existing human biases has been well-documented.[10] And it's not just topics like racism in which the training set may be influenced by human biases: algorithms for finding chemical reaction conditions are influenced by the chemists that program them.[11] This is a particularly fascinating example because few of us commonly think about things like chemical reactions and bias. Chemistry professor Joshua Schrier from Fordham University summarized it well: "Considering machine learning's promise, it's a shame to make an algorithm that's just as stupid as humans because of the way it's trained."[12]

> If human biases can impact machine learning outcomes for designing inorganic reactions, it's something to be cognizant of for other – potentially more consequential – decision-making assisted by AI.

Creative Countermeasures

In the 2017 monograph, *Artificial Intelligence and National Security,* authors Greg Allen and Taniel Chan, identify what they call "Potential Transformative Scenarios."[13] The first of these scenarios is titled "Supercharged surveillance brings about the end of guerilla warfare." Protesters in Hong Kong are innovating and using simple countermeasures to avoid surveillance and identification, such as physical barriers (masks) to lasers to dazzle the facial recognition cameras and wrapping themselves in Mylar emergency blankets to minimize heat (IR, infrared) signatures. These efforts by protestors have been called "[a] war against Chinese artificial intelligence."[14] It suggests that states should not forget about human creativity. Thinking about the nature and how adversaries might employ simple, innovative countermeasures is understudied, if noticed at all.

Conclusions

Reducing the risk from misuse of technology will mean consideration of the highly transnational nature of the critical technology required. Traditional and innovative new approaches to nonproliferation and are important policy elements in reducing the risk of malfeasant applications of technology. Verification still remains a technical as well as diplomatic challenge and the role of international agreements and cooperative programs in the 21st Century is a contested intellectual and policy field.

Science diplomacy has perhaps made the biggest impact in foreign policy as a part of Track II diplomatic efforts: informal diplomacy between individuals who are not officially empowered to act on behalf of the state but are acting in accordance with a state's foreign policy goals and interact through dialogue, as part of increasing cooperation and transparency or in decreasing conflict among states. Track II efforts with nuclear physicists and other scientists during the Cold War are legendary.

Overall, Track II science diplomacy has been an under-utilized tool since the Cold War, which may be ironic considering that technology has enabled the spread, at an unprecedented rate, of scientific knowledge, capabilities, and materials globally. Efforts such as this one organized by CSIS and the PIR Center are critically important.

In the 21st Century, major barriers to effective science diplomacy for national security include three major risks: not being relevant, not being strategic, and not being at the table. The ability t o translate and make relevant the role and importance of science to foreign policy aims is critical. While there are notable exceptions, often this goal is not best accomplished by active research scientists. Similarly, while there are notable exceptions, it is also not often accomplished well by traditional diplomats. In the global information age, there is a critical need for a cohort of individuals who are capable of bridging the divide across technical and national security and foreign policy arenas. In the US, one champion of S&T and foreign policy is institutionalized and embodied in the Science and Technology Advisor to the Secretary of State (STAS).

Technical experts are vital, and lack of expertise can set back efforts by years. The ability to bridge those gaps and work between the technical and the political realms is sometimes over-looked. Once the metaphorical spotlight has been used to illuminate some issue, the science diplomats and other inside and outside the government who possess some mix of technical and policy expertise are responsible for creating, implementing, executing, and assessing the results. It requires empowered and resourced teams of individuals, and increasingly those teams are multi-national, i.e., requiring those with international experience, understanding, and backgrounds.

Much of the concern regarding the potential offensive applications of artificial intelligence is highly speculative and based on worst-case scenarios. The technical and operational veracity of scenarios varies highly from the robust pragmatic realpolitik to Hollywood-like fantasy. Particularly of the industrialized global north, worst-case scenarios garner easy media attention

and can inadvertently drive policy decisions. Choices can be made today, and policy can be implemented in the near future that are likely to shift the balance in favor of maximizing the beneficial and minimizing the negative effects on global security.

Past methods for other technologies that don't take into account the international nature of the science and technology industry are not adequate. Any international regime must be interdisciplinary in focus, cognizant of the multi-polar post-Cold War world, and appreciate the role of private funders, commercial development, and transnational corporations. To be clear, there is much to learn from and leverage in existing arms control and nonproliferation institutions. These starting points and history are valuable; they are not necessarily predictive, however. **Regardless, the challenges in this arena are primarily political rather than technical.**

---

[1] David E Jeremiah (VCJCS, USN, ret), "Nanotechnology and Global Security," Palo Alto, CA; Fourth Foresight Conference on Molecular Nanotechnology, 9 November 1995.

[2] Dmitri Trenin, "Strategic Stability in the Changing World," *Carnegie Endowment for International Peace,* March 2019, https://carnegieendowment.org/files/3-15_Trenin_StrategicStability.pdf; James M. Acton, "Reclaiming Strategic Stability," in Elbridge A. Colby. Michael S. Gerson, Editors, Strategic Stability: Contending Interpretations, Army War College Strategic Studies Institute; Carlisle PA, February 2013, pp 117-146; Adam Stulberg and Lawrence Rubin (editors), *End of Strategic Stability? Nuclear Weapons and Regional Rivalries*, Georgetown University Press, 2018; Pavel Podvig, "The Myth of Strategic Stability," *Bulletin of Atomic Scientists,* October 31, 2012, https://thebulletin.org/2012/10/the-myth-of-strategic-stability.

[3] Elbridge A. Colby. Michael S. Gerson, Editors, Strategic Stability: Contending Interpretations, Army War College Strategic Studies Institute; Carlisle PA, February 2013, https://publications.armywarcollege.edu/pubs/2216.pdf

[4] Charles Perrow, Nor*mal Accidents: Living with High Risk Technologies,* Princeton University Press, 1984; Scott Sagan, The *Limits of Safety: Organizations, Accidents, and Nuclear Weapons,* Princeton University Press, 1995.

[5] Edward Geist and Andrew J. Lohn, *How Might Artificial Intelligence Affect the Risk of Nuclear War?"* RAND, 2018.

[6] Zachary S. Davis, *Artificial Intelligence on the Battlefield,* Center for Global Security Research, Lawrence Livermore National Laboratory, March 2019.

[7] Langelaar, G. C., I. Setyawan, and R. L. Lagendijk. 2000. "Watermarking digital image and video data. A state-of-the-art overview." *IEEE Signal Processing Magazine* 17 (5): 20-46.

[8] Richard, Golden G., and Vassil Roussev. 2006. "Next-generation digital forensics." *Communications of the ACM* 49 (2): 67-80.

[9] Slaughterbots video: https://www.youtube.com/watch?v=HipTO_7mUOw and Paul Schnarre on Why You Shouldn't Fear Slaughterbots https://spectrum.ieee.org/automaton/robotics/military-robots/why-you-shouldnt-fear-slaughterbots

[10] Chouldechova, A., Putnam-Hornstein, E., Benavides-Prado, D., Fialko, O. & Vaithianathan, R. *Proc. Machine Learn. Res.* 81, 134–148 (2018); Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V. & Kalai, A. *Adv. Neural Inf. Proc. Syst.* 2016, 4349–4357 (2016); Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou, "Word embeddings quantify 100 years of gender and ethnic stereotypes," PNAS, 17 April 2018, 115, pp 3635-E3644.

[11] Xiwen Jia, et al., "Anthropogenic biases in chemical reaction data hinder exploratory inorganic synthesis," *Nature,* 11 September 2019, vol 573, pp 251–255, https://www.nature.com/articles/s41586-019-1540-5

[12] Sam Lemonick, "Machine learning can have human bias," *Chemical & Engineering News,* 16 September 2019, vol 97, p 6, https://cen.acs.org/physical-chemistry/computational-chemistry/Machine-learning-human-bias/97/i36

[13] Greg Allen & Taniel Chan, *Artificial Intelligence and National Security,* Harvard Belfer Center, July 2017, p 31.

[14] Via Twitter Alessandra Bocchi @alessabocchi