TECHNO-OPTIMISM, GEOPOLITICS, AND THE FUTURE OF AI

Workshop Summary

January 17 & 18, 2024

Center for Global Security Research
LAWRENCE LIVERMORE NATIONAL LABORATORY

Workshop Summary

Techno-Optimism, Geopolitics, and the Future of AI

Prepared By: Ross Buchanan, Ryan Christenson, Daniel Kroth, Kaitlyn, Lenkeit, Madeleine Lambert, Kimberly Peh, and Brandon Kirk Williams

On January 17 & 18, 2024, the Center for Global Security Research (CGSR) at Lawrence Livermore National Laboratory (LLNL) hosted a workshop on the evolving potential for AI in national security applications, military utility, and civilian economic competitiveness. This event brought together over 50 participants drawn across the policy, military, academic, scientific/technical, and think-tank communities from the United States and a wide spectrum of allied countries. The discussion was guided by the following objectives:

- To bring together technology developers and security experts to develop improved shared understanding of opportunities and challenges
- To understand how AI might shape the future security environment
- To understand how the future security environment might shape AI

Key Conclusions

1. As in most workshops on AI, the opening comments centered on the remarkable pace of progress in the field. AI is playing an increasing role across numerous national security vectors, such as stockpile stewardship, biosecurity, and advanced manufacturing. Large language models (LLMs) are experiencing substantial changes in modality as they move from text to images, measurements, and complex science. Foundation models can be put to more uses, both positive and negative. If scientific discovery was seen to alternate between periods of normal science and paradigm shifts, the sense of the group was that the last few years has seen the latter in this field.

2. There is some debate, however, whether the pace of this rapid advancement is sustainable as we look out to the late 2020s and early 2030s, even if there is consensus that there is still room to grow. Few states and companies have the resources necessary to compete. The computing power, known as compute, necessary to drive LLMs is expensive, limited, and environmentally problematic. Training costs for increasingly complex LLMs are increasing. Educating, attracting, and retaining human talent is difficult. It remains an open question whether innovation will drive costs down and diversify high-end capabilities, or if the barriers to entry will mean that only a few have access to cutting edge models in this field. Much of this depends on whether the market for AI becomes oversaturated or undersaturated in a traditional business sense.

3. Although AI is a technology in which the United States is in competition with both Russia and China, the dimensions of that competition are very different with each country. Russia's approach is more evolutionary than revolutionary. Russia is behind the United States and China, but it has shown the capability to innovate and adjust under fire in Ukraine. Russian

military challenges have traditionally not been on conceptualization but true integration of capabilities into military operations. Many Russian military challenges regarding AI are shared by the United States. A key question with Russia is whether being the AI second mover means it is eternally disadvantaged on the battlefield or able to gain many of the advantages of United States or Chinese innovation at much less cost.

4. China in contrast has a clear goal of "leapfrogging" the U.S. and maintaining a clear decisive advantage in AI. As AI technology is assessed to provide a "head wild goose effect," their leadership is 100 percent committed to this goal and is devoting massive resources across the board. China manages innovation differently than the United States, throwing money widely to see what emerges. Chinese leadership in the Chinese Communist Party (CCP) can also achieve great effects in ordering adoption of AI tools, such as the use of WeChat pay. AI is seen to be key to urban population management (i.e. sewage, transit, power, traffic, etc.) and thus CCP internal domestic control. The People's Liberation Army (PLA) has built AI into its goals of mechanization, informatization, and intelligentization regarding both capabilities and operations. Chinese military officials see a clear advantage to be gained from AI tools, both in terms of synthesizing information across the battlespace and coordinating complex operations with speed and precision.

5. Three key themes in the discussion of U.S.-Russia-China dynamics throughout the workshop were: 1) the problem with the metrics used to assess AI competition, 2) the difficulty in evaluating the pros and cons of each systems' government innovation models, and 3) the lack of assessment methodologies. The most often cited quantitative metrics of competition – dollars spent, articles published, patents developed, PhDs graduated, market share, and so on – are all seen as potentially misleading indicators of being "ahead" or "behind." Numbers do not necessarily equal quality, and regimes like the CCP have a vested interest in attempting to inflate their successes and capabilities. Similarly, regime types have different advantages and disadvantages. More authoritarian regimes can spend huge resources and dictate directions, but more democratic regimes can move faster through public-private partnerships and more open innovation ecosystems. Advanced methodologies like net assessment and wargaming are hindered by the difficulty with metrics. One solution could be to incorporate better understanding of business approaches.

6. The China-Russia strategic partnership poses some interesting challenges for innovation and competition. With China as a clear leader in AI technology, a question was what it could gain from Russia. Russia possesses two assets of clear potential value. It has a trove of military-related training data from its war in Ukraine, which can be used to train and improve Chinese models and system, and in combat experience in the Middle East. Russia also has extensively studied the fissures in U.S. society, which can be used by the Chinese to plan for disinformation campaigns. Russia in turn can be a faster second mover in this field, keeping a close enough pace with the frontrunners to be dangerous at a fraction of the cost to develop and design the computing power and models themselves.

7. It is a truism that AI technologies are rapidly becoming available to a host of new actors. The participants noted that some useful methodologies can be applied to understanding this technology proliferation. Outside of the leading states, countries such as Iran, North Korea,

and Turkey have shown impressive capabilities integrating technology into weapons like armed drones. The lines between high tech and low tech in this space are hard to determine. Types of non-state actors include major multinational corporations, smaller startups, international criminal organizations, paramilitary or terrorist groups, transnational organizations, and even individuals. The demonstrated capabilities and limitation of private sector technology actors in the war in Ukraine was a notable theme.

8. Analysis of emerging actors and threats often overlooks important technical details necessary to understand capabilities and threats, leading to overhyped threats of AI + X group + Y adjacent technology. Consideration must be paid to the technology or tool itself, the concept of potential operations, and the method of employment. There are different types of compute, different ways to access data, and different algorithms. Integrating and testing algorithms against the data is difficult. Information can be purchased or acquired, but it may not be correct or properly tailored for a specific use. Hardware still matters. One needs operators, who must be trained, organized, and directed. All analysts need to be somewhat AI literate to understand the technical capabilities and limitations related to their particular field of study.

9. One participant argued that there is "a massive imperative" to get the cooperation piece right between the United States and its allies and partners. Nations such as Australia have incorporated AI into their key national security and defense strategies are thinking through how to leverage AI to best effect given their societies and situations. Organizations like NATO have articulated an ambitious vision for transformation through new partnerships and innovation accelerators. Initiatives like AUKUS provide enormous opportunities for technology sharing and innovation. The workshop highlighted the breadth of the cooperation required in this space, including scientific cooperation, regulatory cooperation, privacy enhancement, export controls, rules/norms/standards setting, incident reporting, election monitoring, and response coordination.

10. Much of this has to do with framing the international competition on AI as something other than a bilateral competition between the United States and China, between democratic and authoritarian regimes, or between leading states and the so-called "Global South." The global effort, even if in points led by the United States, is a multilateral effort. States such as Singapore are important partners but may not fit cleanly into democratic/authoritarian columns. Europeans are focused on risk management, but many states in the Global South have different goals and perspectives. They may see AI-technologies as critical for achieving their economic modernization goals. They may see Chinese technologies as coming with fewer strings and better aligned with their political goals. They may not share the techno-pessimism of the United States and its allies or many U.S.-based multinational corporations.

11. Global AI governance was depicted as "an intense competition over who sets the norms." Likeminded states have come together in a wide variety of venues to set norms and reach consensus on AI risks and safety. This includes the G7, AI Partnership for Defense, the Council for Europe, OECD, the UK AI Safety Summit, and work in the UN General Assembly. A great deal of work has been completed, but this remains the early stages of attempting to regulate effectively. Many recognized the imperative of regulating early. The United States has a clear goal in preserving its competitive edge through constraining potential

adversaries, promoting domestic AI innovation, attracting talent, and strengthening relationships with allies and partners. Governance is forced to take on many functions, including mitigation, oversight, incident reporting, incentivization, and regulation.

12. Efforts at governance come into direct conflict with the dynamics and perceived benefits of racing, technology proliferation, and diffusion. There are benefits of incorporating AI into nuclear-related systems, such as improved speed and efficiency, but also well understood stability risks surrounding autonomy, entanglement, and misinformation. Biosecurity was the most frequently cited area of concern during the workshop for AI-enabled advancements, with an already low barrier to entry made seemingly even lower. Governance also often runs into problems related to societal values, partisan divisions, and differing sets of ethics.

13. A key question for AI governance remained where to do it to best effect, which requires understanding the technical details relevant to the field. The intersections between AI and other domains may be best dealt with using regulatory mechanisms already in place for other domains, such as adapting the Biological Weapons Convention to address AI issues. Participants also emphasis that the need to focus on particular AI-specific effects, and not set standards that did not solve the problems at hand or set the bar too high or too low to be effective.

14. Workshop participants emphasized the necessity of monitoring AI's development. ChatGPT's introduction in November 2022 and the rising utility of LLMs are forcing a recalibration of previous conceptions of AI applications for military use. It is premature to identify LLMs' immediate use-cases for national security missions, but LLMs and frontier models possess near-term potential for advancing science and technology. LLMs suffer from a variety of technical hiccups—such as widely reported phenomenon dubbed hallucinations—and cybersecurity questions via prompt injection attacks or data poisoning from malicious actors that will prevent the U.S. government from integrating models into national security missions.

Panel 1: Calibrating the AI Future

- Along what vectors are technologies likely to develop?
- How useful is the last decade as a guide to the next decade?
- What future milestones can be predicted?

The past few years have witnessed an AI explosion that outpaced expectations where advances are occurring years before expected development and implementation. The deep learning revolution after 2010 applied powerful and precise machine learning (ML) techniques to increasingly complex tasks performed by LLMs and foundation models. Public awareness of LLMs for AI, however, did not occur until OpenAI's ChatGPT's 2022 release that ushered Generative AI into the public consciousness. Since, OpenAI and competitors have released Generative AI products to varying degrees of adoption. Models will likely improve and accelerate research and development in a wide range of national security and private sector applications. For instance, some experts predict that LLMs will transform scientific discovery and synthesis soon. Even though predicting milestones will prove difficult, panelists agreed that AI will mature into a ubiquitous general-purpose technology sooner than initially anticipated.

The current period of rapid growth may continue into the 2030s before slowing. If the pace of progress proceeds as expected, AI in the early 2030s will be vastly more powerful than today. Nevertheless, constraints exist. Future AI models will face hurdles from compute and resources. Compute has doubled every three and a half months and shows no signs of slowing. Increasing the efficiency of compute will eventually collide with chips approaching the smallest nanometer size. A consequence of this limitation is that running data-hungry AI models is becoming increasingly energy-intensive and expensive. Specialized computing systems can partially mitigate this limitation, but the development and construction of such systems at scale is costly. OpenAI's GPT-4 reportedly cost $100 million to train. Given these limitations, further development of models may depend on the models attaining greater efficiency. Panelists concluded that nations and a handful of companies will likely preserve control over cutting-edge AI as a result.

The proliferation of open-source AI models in the coming decade will demand flexible understanding for AI leadership, use, and governance. For instance, a variety of nations and nonstate actors may use open-source models to stay competitive with the United States. Open-source models can be deployed, modified, and repurposed by any user. Closed-source models will remain proprietary, but training these models will pose a significant challenge due to the data-intensive and therefore energy-intensive nature of the process. The compute infrastructure required to train and ultimately deploy a frontier model is also substantial and often requires specialized hardware. Nevertheless, a nation or company that owns a model or hardware infrastructure cannot be guaranteed to remain permanently ahead as open-source models advance. Policymakers and the private sector must monitor the proliferation of open-source models that may allow adversaries to fast follow.

Calibrating the AI future must also encompass solutions for incorporating robust safeguards before open- and closed-source models are accessible. Adversaries could conceivably use open-source models for malicious actions, such as facilitating cyber-attacks or developing

sophisticated information campaigns. There are currently no robust ways to incorporate safeguards into open-source models; any user who has access to the model itself can easily disable any safeguards. Although companies such as OpenAI and Anthropic red team and align closed-source models, the models are vulnerable to attacks and users can jailbreak the models to generate data that many governments would consider harmful. Future models may be able to incorporate guardrails at a more fundamental level. There is debate, however, if companies or nations can deploy these technical safeguards feasibly to keep pace with the staggering pace of development.

Panel 2: Russia, China, and the National Security Applications of AI

- What are their objectives and level of effort?
- What progress have they made and can they be expected to make?
- What strengths and weaknesses shape their efforts?

Panel 2 took stock of Russia's and China's levels of effort in harnessing AI to achieve national security objectives and the factors that may facilitate or hinder their respective successes. Russia's evolutionary stance on AI differs from China's and the United States' perception of AI through a revolutionary lens. As such, Russia is not seeking such technological breakthroughs or AI adoption levels that would position it as a competitor that rivals China's competitive relationship to the United States. Russia's efforts in Ukraine demonstrate its lack of AI integration given its limited success in many of its flagship efforts, such as, the use of uncrewed vehicles, augmenting command and control, and information and cyber operations.

In contrast, China is fully committed to AI development, and it seeks to establish itself a global leader in this domain by 2030. In this competition to become a global leader, China benefits from its ability to quickly adopt technology and at scale, which is a product of its top-down structure. Policy-wise, China has adopted a strategy of military-civil fusion that allows its military to capitalize on significant civil investments in AI. The PLA is motivated to integrate AI into its military because it places information control at the core of its operational concept. In the PLA's strategic doctrine, warfare based on information control would rest on a nation's ability to quickly process and use information to improve decision-making. AI and autonomous platforms are thus important because they move warfare into an intelligent age where unmanned systems and human-machine teaming are envisioned to enhance decision-making processes and overcome threats of human or machine errors in combat.

Determining progress or leadership in AI, however, is challenging and cannot be easily measured. Panelists agreed that progress and technological advancements cannot be understood in a vacuum. Much of a nation's progress depends on contextual factors like bureaucratic culture, corruption, budgets and procurement, and institutional competence. The United States, China, and Russia all suffer from a shortage of workforce and talent, and all three countries continue to struggle in properly and effectively integrating AI for the purpose of achieving broader national security goals. Each nation's military is searching for the best AI systems to integrate operational and battlefield data in spite of data limitations. China, furthermore, lacks training data and information on the U.S. population that are vital for

disinformation campaigns to influence U.S. society. Training LLMs in China may be hampered by CCP censorship policies, which has led Chinese companies to restrict Generative AI outputs.

Competing with Russia and China over AI may unfold asymmetrically and demand managing allied relationships. U.S. policymakers and the analytical community should monitor Russian and Chinese data sharing and AI collaboration. Thus far learning and collaboration remains limited. A more pressing concern lies in countries developing AI hardware and systems that may also be aiding both Russia and China by supplying investment capital, machinery, and know-how. To the extent that these countries are also assisting Beijing and Moscow, China and Russia will continue to make progress despite U.S. export control policies to complicate their efforts. Policymakers should prepare for a new style of diplomacy with partners and allies to manage AI and technology competition.

Panel 3: AI and the "Democratization" of Conflict

- To what extent will AI empower new actors, both state and non-state?
- What new forms of conflict might emerge?
- What impact might this have over the next decade?

Both states and non-state actors may benefit from a democratization of AI-powered conflict, but obstacles will remain to initiating new forms of conflict. Actors may leverage AI along three axes: compute, algorithms, and data. For example, actors can use edge computing for dynamic jamming, buy algorithms and utilize them for deception, and use data to sustain information campaigns. In each case, however, significant limitations arise. AI surveillance systems rely on high levels of compute but have middling precision rates when deployed. Algorithms are imperfect and may not perform to expectations in dynamic conflict settings. Similarly, despite large volumes of available data, training data for targeting is limited. Panelists called for balanced and nuanced analysis because new actors may struggle if deploying AI systems that are ill-suited for the challenge of warfighting.

Non-state actors may have the ability to acquire advanced technologies and possess the motivation to use AI for malicious purposes. Non-state actors include companies, international criminal groups and cartels, state-backed non-state actors, organizations, and individuals. When evaluating non-state actors' use of AI, it is important to consider which actors may possess the will and resources to use AI-enabled tools. Many actors may struggle to integrate AI to meet objectives based on the resources required. Technological and financial constraints may thus limit the utility of AI in the next decade. A better understanding of these factors will assist in horizon scanning, multilateral AI arms control, and the conditions under which AI systems would be deployed at scale by state and non-stake actors alike.

Panelists agreed that a focus on advanced nations overlooks the ascendance of small and middle-power states achieving second mover advantages. The discussion around AI-enabled conflict often revolves around advanced nations due to the digital divide. Nevertheless, several states have acquired and demonstrated the capability to quickly adopt AI and other advanced technologies. For instance, Iran is now a key exporter of drones; Turkey is a leading nation in the

building of unmanned aerial vehicles; and North Korea has shown, in the early 2000s, its ability to clone Windows software despite the lack of contact with Microsoft or the West. Each country's successes highlight the possible existence of a second mover advantage and the speed at which emerging technologies can mature, diffuse, and be adopted by actors outside of advanced nations.

The difficulty lies in keeping these actors accountable given the diversity of actors and the ease of proliferation. Many of these actors exist outside of regulatory regimes and non-state actors face less pressure in acceding to international regimes. Different countries, too, have divergent philosophies towards regulation, and it is unclear how arms control will be practiced within countries and globally. The increased diversity of actors may also create new opportunities in areas like conflict management and responsible use of military AI. The United States-backed Political Declaration of Responsible Military Use of Artificial Intelligence and Autonomy is a normative framework that will build consensus on reducing risk and ensuring safe AI deployment. Additionally, in the future, governments may leverage public-private partnerships to convince companies and other stakeholders to advocate for legally binding AI regulation.

Panel 4: The United States, its Allies, and the National Security Applications of AI

- What are their objectives and level of effort?
- What progress have they made and can they be expected to make?
- What strengths and weaknesses shape their efforts?

The United States and its allies recognize AI's transformative national security potential, and all parties are ramping up cooperative initiatives to meet the challenge of AI alignment. Although AI adoption disparities could destabilize alliance relationships, European and Pacific allies are coordinating with the United States to improve AI readiness while also shifting domestic resources to keep pace with AI development. All parties must overcome cultural, regulatory, and technological differences to meet this objective, and a considerable amount of work looms. Fortunately, the United States and its allies can leverage long-standing alliance relationships to arrive at political and technological solutions.

Thus far, the United States' allies in Europe and the Pacific are acting promptly to prepare for AI's national security disruptions. Allied capitals are coalescing on the necessity of setting export controls to protect domestic industries and prevent adversaries from capitalizing on lax oversight. Progress on this front is matched with diplomatic backing for agreements such as the Bletchley Declaration as well as the Political Declaration on Responsible Military Use of AI and Autonomy. These agreements are establishing the conditions for setting norms and standards in the United Nations and standard-setting bodies. Domestically, allies and the United States launched initiatives to shape private sector research and development for national security applications. NATO's Defence Innovation Accelerator for the North Atlantic, Australia's Advanced Strategic Capabilities Accelerator, and the Department of Defense's Office of Strategic

Capital are focusing on delivering near-term operational tools. Overall, the United States and allies made quick progress on multiple fronts to prepare for a global AI transformation.

Despite this progress, strengths and weaknesses characterize each country's efforts to keep pace with AI. NATO and Pacific allies are aligned on the necessity to move fast. This consensus is inspiring a new culture of experimentation to integrate interoperable tools and initiate new multilateral and bilateral groups like the U.S.-EU Trade and Technology Council to sharpen protective toolkits. The United States and its allies are also readying their private sector technology partners to help weave AI into national security and basic government functions.

Regardless, the speed of the United States' private sector and classified dimensions of programs such as the Replicator Initiative may strain alliance relationships. Few countries or even companies can rival the dynamism of the United States' AI sector. The United States will remain the fastest mover, and allies' struggles to catch up may allow technologically advanced rivals like China to prevent allied alignment. Additionally, the Department of Defense's Replicator Initiative is preparing to operationalize AI and autonomous vehicles, but few allies can seamlessly join due to classification and regulatory constraints. Cold War agreements such as the Wassenaar Agreement and the International Traffic in Arms Regulations are injecting friction into U.S.-allied AI interoperability. With time, these issues may be overcome as nations' defense industrial bases, AI ecosystems, and regulatory regimes adapt to meet the new national security environment and the accelerating dynamics of AI competition.

Panel 5: The Dynamics of AI Competition

- What are the potential benefits, costs, and risks of competition?
- How much competition is "enough?"
    - Can the U.S. and its allies safely choose not to compete in certain ways?
- Is winning possible?
    - Are there strategic advantages in AI that can be enduring?
- How can we know who is winning?
    - How can net assessment methodologies be adapted to this challenge?

Since 2022, the accelerating tempo of AI development and semiconductor export controls elevated AI competition to new heights. A fusion of breakthroughs and policies to protect national competitiveness are unleashing new competitive dynamics that will likely accelerate as AI diffuses across nations and militaries. It is premature to determine which nation will win or lose AI. In fact, a win or lose framing may confuse more than clarify. Many states will search for second mover advantages that limit costs but allow for quick adoption to remain competitive. Thus it will be difficult to determine which nations sustain the best competitive advantage and nurture the talent to drive innovation. Competing for talent will only amplify as nations and companies vie for the best-trained individuals. U.S. legacy technology companies and startups will face increasing demand for global talent that may shape leadership. When comparing countries, business data may illuminate more than metrics such as publications, advanced degrees, or patents. Many of those data points may cloud the picture of global leadership and

produce poor strategic outcomes. For the United States, the best competitive strategy remains to influence standards and norms as well as communicate that the world's superior AI ecosystem resides in the United States.

U.S. policymakers are keenly aware that monopolizing AI is not only futile, but also counterproductive to ensuring responsible AI use. Global AI developers such as France's Mistral will release open-source LLMS, and the United States has little ability to control open-source model proliferation. For the United States, competing requires leadership in establishing AI norms, standards, and guardrails in coordination with allies and partners. This strategy will require deft diplomacy. Many nations may not embrace the United States' vision of responsible use of AI due to differing political systems and ideologies. Areas for potential agreement with diverse global stakeholders include watermarking and copyrighting data. Those technological solutions are not a panacea for the challenges confronting democracies. The United States cannot compete comprehensively over AI, but enduring strategic advantages will emerge from setting iterative global standards that adapt to AI's disruption.

At a broad level, adversaries are adapting to a post-ChatGPT moment by seeking second mover advantages that will sustain competitiveness and circumvent U.S. export controls. Officially sanctioned Chinese ChatGPT clones like Baidu's Ernie Bot quickly reached a performance ceiling and Chinese regulation threatened to stymie LLM development. To overcome these constraints, the CCP is warming to open-source models with an iterative regulatory regime because the CCP perceives open-source model as an asymmetric advantage for preserving China's competitive stature and bypassing export controls. For Russia, brain drain, export controls, and limited global partnerships leaves Moscow with few options other than to seek second mover advantages wherever possible.

Panel 6: AI Competition and Nuclear Stability

- What are the risks?
- Can they be mitigated?

Panelists concluded that AI competition and AI use in nuclear command, control, and communication poses little risk to stability, but it will alter the nuclear landscape. Panelists emphasized that current AI integration in the nuclear decision-making process does not include autonomous nuclear launch capabilities. Deeper integration into the decision support system is certainly possible, but there are risks. AI systems can have biases and other flaws that are not immediately apparent but can undermine effective decision-making. Better understanding AI systems, biases, and other imperfections is necessary to effectively employ them in decision-making. The risk equation involving threat vulnerability, and consequence is highly uncertain in the AI context. Differences in defense decision-making approaches are evident among major powers. Russia relies more on mathematical formulas and focuses on certainty, while China's deterrence strategy entangles nuclear and conventional command and control to disrupt U.S. deterrence calculations.

AI is likely to play a critical role in strategic decision-making processes. AI's ability to quickly sort and synthesize large volumes of information from diverse sources can improve decision makers' situational awareness. AI may be employed for low-risk applications such as improving early warning systems, collating sensor data, and managing predictive maintenance and logistics. Before AI can be integrated into the decision-making process, trust must increase. Systems must also be hardened against cyber-attacks and misinformation. LLMs may unleash a new era of misinformation that can influence leaders' calculus that could threaten stability.

AI may enhance the ability to break one or more links in a kill chain with considerable effects for nuclear stability. Kill chain robustness determines how effectively one can target an enemy's nuclear deterrent, and how survivable one's own nuclear deterrent is against an enemy first strike. AI will upset nuclear stability if it greatly decreases the survivability of a nuclear deterrent to a first strike. AI is likely to improve abilities to both target nuclear assets and to better hide nuclear assets. One plausible future is an AI arms race expressed primarily through the development of decoy and counter-decoy abilities. For instance, decoys can be used to conceal the location of military assets from adversary surveillance and reconnaissance. AI can enhance the effectiveness of decoys that allows the decoys to broadcast emissions that convincingly mimic actual military assets. For instance, a submerged decoy could be used to mimic a submarine's acoustic signature to distract adversary anti-submarine warfare assets.

AI integration is altering the landscape of nuclear stability and conventional strikes. AI's impact on nuclear arms competition is multifaceted, affecting both short and long-term dynamics. For instance, Russian defense analysts are deeply engaged in understanding the potential enhancements AI might bring to U.S. capabilities, particularly concerning their nuclear arsenal's vulnerability to U.S. conventional counterstrikes. This focus on AI extends to the survivability of second-strike forces, with AI potentially linking or disrupting kill chains through advanced countermeasures. AI-enabled underwater sensors and Unmanned Underwater Vehicles may increase ocean transparency, but physical limitations in signal transmission through saltwater remains a significant challenge.

Panel 7: The Future of AI Governance

- What alternative models should be considered, both formal and informal?
- Which are feasible and which not? Why?

AI poses a unique governance challenge for the United States and its allies. It is ubiquitous, complex, and difficult for the layperson, policymaker, or regulator to fully understand. Beyond this, the extraordinary pace of AI development and the technology's ability to permeate regulatory boundaries complicate governance. Nonetheless, it is important to act now. As AI and the enabling technologies develop, highly capable systems will be increasingly accessible, promising both incredible advances and potentially significant harms. Likely near-term advances in robotics and related technologies will allow AI to move to real world capabilities which may be more difficult to control. It is therefore imperative for nations to regulate AI, even frontier AI, and cooperate on global governance measures.

Governing AI calls for novel regulation, adapting existing regulatory frameworks, and expanding current regulatory authorities. An “AI retooling” is often overlooked in discussions of AI regulation. Nonetheless, it shows significant promise. For example, preventing LLMs from assisting a person in developing a chemical or biological weapon may prove difficult because the process is not theoretically complicated and relies on freely available information. However, in the process of synthesizing a pathogen or deploying a chemical agent based on LLM-supplied data, a malicious actor is likely to commit noticeable and prosecutable offenses. This is also true at the international level, where regulations on wartime behavior will still apply to AI-based weapons. Empowering relevant regulators, domestic and international, to enforce current law in an AI-enabled world is an impactful and feasible first step.

In addition to this retooling, new governance mechanisms are needed to keep pace with AI evolution. This will include AI committees in key international forums like the G7 or the Council of Europe. These bodies could promote auditing or licensing regimes, reasonable restrictions, incident reporting, or other prudent measures. Nuanced and tailored domestic strategies are also likely to be necessary. These approaches are likely to require some form of disclosure, licensing, certification of provenance, or auditing. Governments and technologists are debating the merits of open-source models such as Meta’s Llama 2 or platforms such as Hugging Face that share open-source tools. Some parties argue that adversaries or nonstate actors may abuse open-source tools. Others insist that sufficiently powerful models for many forms of malicious activity will be available and that prohibiting open sourcing could hamper innovation and impede the development of responsible regulatory or technological countermeasures against malign use.

Panelists concluded that governments should adopt agile and iterative measures for AI governance. This will allow countries to respond to the potentially rapid and nonlinear development of AI technologies and to conform to ethical and normative preferences in varied cultural settings. Normative frameworks are an important precursor to formal international agreement and can shape behaviors. While ethics may differ across cultures, large areas of commonality can form the basis for legitimate governance. The United States and its allies in Europe and Asia can develop then implement transparent governance regimes that can be fairly applied globally. Inclusive governance will require, nevertheless, incorporating the perspectives and priorities of Global South nations and peoples.

Center for Global Security Research
Lawrence Livermore National Laboratory
P.O. Box 808, L-189 Livermore, California 94551
https://CGSR.llnl.gov

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-MI-861717