BROOKINGS

TechTank

It is time to negotiate global treaties on artificial intelligence

John R. Allen and Darrell M. West Wednesday, March 24, 2021

The U.S. National Security Commission on Artificial Intelligence recently made the news when its members warned that America faces a national security crisis due to insufficient investment in artificial intelligence and emerging technologies. Commission Vice Chair Robert Work argued "we don't feel this is the time for incremental budgets ... This will be expensive and requires significant change in the mindset at the national, and agency, and Cabinet levels." Commission Chair Eric Schmidt extended those worries by saying "China is catching the US" and "competition with China will increase."

This is not the first time the country has worried about the economic and national security ramifications of new technologies. In the aftermath of World War II, the United States, Soviet Union, China, France, Germany, Japan, the United Kingdom, and others were concerned about the risk of war and the ethical aspects of nuclear weapons, chemical agents, and biological warfare. Despite vastly different worldviews, national interests, and systems of government, their leaders reached a number of agreements and treaties to constrain certain behaviors, and define the rules of war. There were treaties regarding nuclear arms control, conventional weapons, biological and chemical weapons, outer space, landmines, civilian protection, and the humane treatment of POWs.

The goal through these agreements was to provide greater stability and predictability in international affairs, introduce widely-held humanitarian and ethical norms into the conduct of war, and reduce the risks of misunderstandings that might spark unintended conflict or uncontrollable escalation. By talking with adversaries and negotiating agreements, the hope was that the world could avoid the tragedies of large-scale conflagrations, now with unimaginably destructive weapons, that might cost millions of lives and disrupt the entire globe.

With the rise of artificial intelligence, supercomputing, and data analytics, the world today is at a crucial turning point in the national security and the conduct of war. Sometimes known as the AI triad, these characteristics and other weapons systems, such as hypersonics, are accelerating both the speed with which warfare is waged, and the speed with which warfare can escalate. Called "hyperwar" by Amir Husain and one of us (John R. Allen), this new form of warfare will feature levels of autonomy, including the potential for lethal autonomous weapons without humans being in the loop on decision-making.

It will affect both the nature and character of war and usher in new risks for humanity. As noted in our recent AI book "Turning Point," this emerging reality could feature swarms of drones that may overwhelm aircraft carriers, cyberattacks on critical infrastructure, AI-guided nuclear weapons, and hypersonic missiles that automatically launch when satellite sensors detect ominous actions by adversaries. It may seem to be a dystopian future, but some of these capabilities are with us now. And to be clear, both of us, and more broadly the world's liberal democracies, are struggling with the moral and ethical implications of fully autonomous, lethal weapon systems.

In this high-risk era, it is now time to negotiate global agreements governing the conduct of war during the early adoption and adaptation of AI and emerging technologies to the waging of war and to specific systems and weapons. It will be much easier to do this before AI capabilities are fully fielded and embedded in military planning. Similar to earlier treaties on nuclear, biological, and chemical weapons in the post-war period, these agreements should focus on several key principles:

- Incorporate ethical principles such as human rights, accountability, and civilian protection in AI-based military decisions. Policymakers should ensure there is no race to the bottom that allows technology to dictate military applications as opposed to basic human values.
- Keep humans in the loop with autonomous weapons systems. It is vital that people make the ultimate decisions on missile launches, drone attacks, and large-scale military actions. Good judgment and wisdom cannot be automated and AI cannot incorporate necessary ethical principles into its assessments.

- Adopt a norm of not having AI algorithms within nuclear operational command and control systems. The risk of global destruction is high with AI-based launch on warning systems. Since we do not know, and may never know, exactly how AI learns from training data, it is important not to deploy systems that could create an existential threat to humanity.
- Protect critical infrastructure by having countries agree not to steal vital commercial data or disrupt power grids, broadband networks, financial networks, or medical facilities on an unprovoked basis through conventional digital attacks or AI-powered cyber-weapons. Creating a common definition on what constitutes critical infrastructure will be important to the implementation of this principle.
- Improve transparency on the safety of AI-based weapons systems. It is crucial to have more information on software testing and evaluation that can reassure the public and reduce the risks of misperceptions regarding AI applications. That would provide greater predictability and stability in weapons development.
- Develop effective oversight mechanisms to ensure compliance with international agreements. This would include expert convenings, technical assistance, information exchanges, and periodic site inspections designed to verify compliance.

The good news is there are some international entities that already are working on these issues. For example, the Global Partnership on Artificial Intelligence is a group of more than a dozen democratic nations that have agreed to "support the responsible and human-centric development and use of AI in a manner consistent with human rights, fundamental freedoms, and our shared democratic values." This community of democracies is run by the Organization for Economic Cooperation and Development and features high-level convenings, research, and technical assistance.

That said, there are increasingly calls for the technologically advanced democracies to come together to aggregate their capacities, as well as leveraging their accumulated moral strength, to create the norms and ethical behaviors essential to governing the applications of AI and other technologies. Creating a reservoir of humanitarian commitment among the democracies will be vital to negotiating from a position of moral strength with the Chinese, Russians, and other authoritarian states whose views on the future of AI vary dramatically from ours.

In addition, the North Atlantic Treaty Organization, European Union, and other regional security alliances are undertaking consultations designed to create agreed-to norms and policies on AI and other new technologies. This includes effort to design ethical principles for AI that govern algorithmic development and deployment and provide guardrails for economic and military actions. For these agreements to be fully implemented though, they will need to have the active participation and support of China and Russia as well as other relevant states. For just as it was during the Cold War, logic should dictate that potential adversaries be at the negotiating table in the fashioning of these agreements. Otherwise, democratic countries will end up in a situation where they are self-constrained but adversaries are not.

It is essential for national leaders to build on international efforts and make sure key principles are incorporated into contemporary agreements. We need to reach treaties with allies and adversaries that provide reliable guidance for the use of technology in warfare, create rules on what is humane and morally acceptable, outline military conduct that is unacceptable, ensure effective compliance, and take steps that protect humanity. We are rapidly reaching the point where failure to take the necessary steps will render our societies unacceptably vulnerable, and subject the world to the Cold War specter of constant risk and the potential for unthinkable destruction. As advocated by the members of the National Security Commission, it is time for serious action regarding the future of AI. The stakes are too high otherwise.