CNA
ANALYSIS & SOLUTIONS

AI AND AUTONOMY IN WAR: UNDERSTANDING AND MITIGATING RISKS

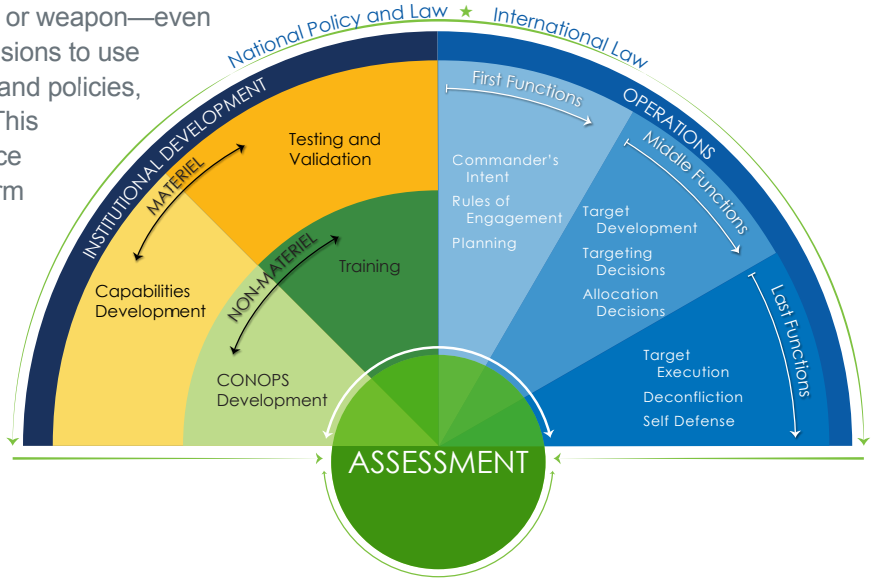
Significant advances in artificial intelligence over the past decade have changed our way of life, and the impacts of AI are only expected to accelerate. AI is increasingly discussed as a source of good in many areas, such as medicine, education, and law enforcement. At the same time, the idea of military applications of AI and the related attribute of autonomy has created considerable controversy. There are strong concerns about these technologies, even speculation that they could lead to the end of the world. These concerns raise important questions:

- How do the actual risks of weaponizing this technology compare to those commonly discussed?
- Are states and the international community effectively managing these risks?

A new report from the Center for Autonomy and AI at CNA, *AI and Autonomy in War: Understanding and Mitigating Risks*, examines commonly held concerns as reported in the media or voiced in international gatherings. We find that the premises for some concerns about AI and autonomy in war are out of step with the current state of the technology, while others do not consider the way military systems are actually used as part of a larger process for delivering force. Several of these warnings assume the possibility of "general AI," what Elon Musk called "god-like superintelligence" in his own critique of military AI. But evidence suggests that general AI is still many decades away—if it is possible at all.

Current applications are all examples of a different type of AI, narrow AI, in which machines perform specific, pre-programmed tasks for specific purposes. These applications of narrow AI to war do carry concerns and risks—including civilian casualties and fratricide—but they are fundamentally different from those associated with general AI. Our report identifies factors associated with the current and near-future state of the technology that could introduce operational risk if not mitigated. Many of the serious risks we consider—including overconfidence by human operators of the technology and the inability of current test procedures to evaluate AI systems—are not commonly recognized. These risk factors can be blind spots for militaries, which tend to focus on developing a capability without considering the set of enablers necessary for its effective use in a way that reduces operational risks.

We analyze risk-mitigation opportunities by drawing on decades of military practice, especially recent battlefield lessons from Iraq and Afghanistan. The analysis is grounded in the understanding that there is no such thing as an autonomous soldier or weapon—even if it is a so-called "autonomous" weapon. All decisions to use force are made in a larger context of processes and policies, which we illustrate in the accompanying figure. This framework for human control over the use of force informs our risk-mitigation analysis and can inform international and domestic discussions about AI and autonomy in war.



Finally, we note that AI and autonomy also provide opportunities for improving humanitarian outcomes in war. This technology should be seen objectively, in light of potential benefits as well as risks, so that nations can seek to both leverage the benefits and mitigate the risks to improve the conduct of war.

RECOMMENDATIONS

We offer a number of recommendations for mitigating risks from the use of AI and autonomy in war. The first set is for nations considering the military implementation of those technologies, to enable them to better address clear and present risks. Recognizing the need for additional and productive discussion regarding AI and autonomy in war, the second set addresses needed dialogues to consider the risks of those technologies and how to mitigate them.

For countries considering the military use of AI and autonomy:

- Address risk factors impacting operational safety—including operational considerations, institutional development, and law and policy—in order to both improve effectiveness and promote safety.
- Consider policies to address the risk that AI could increase the opacity of targeting decisions. Policies should exclude the use of AI in "signature strike" scenarios, in which targets are not specifically identified, but bear the signature of combatant activity.
- In addition to mitigating risk factors, look for opportunities to use the benefits of AI and autonomy to improve the conduct of war.

For needed dialogues on the risks of AI and autonomy in war:

- Separate out the two cases of general and narrow AI, since the two are distinct, with very different sets of risks and different timelines for development.
- Hold deliberate, inclusive debates concerning AI and autonomy in war, requiring arguments to be supported with reason and evidence and allowing different views to be fairly exchanged.
- Discuss the risk that AI will increase the opacity of targeting decisions and the steps that can be taken to avoid this.
- Consider risk factors identified in this report as a way to frame discussions on how to pursue the safety of AI and autonomy in war. Those discussions should include operational considerations, institutional development, and law and policy.
- Consider potential opportunities for using AI and autonomy to improve the conduct of war.

CNA CENTER FOR AUTONOMY AND AI

CNA's Center for Autonomy and AI supports the U.S. goal of effectively incorporating autonomy, AI, and related technologies into military capabilities. Autonomy and AI represent revolutionary technologies in warfare which offer opportunities to the U.S. for countering and deterring emerging threats, addressing security challenges and advancing U.S. national interests. But this opportunity is by no means certain, since autonomy also offers potential asymmetric advantages to near-peer competitors, some of which have been pursuing these capabilities aggressively. Likewise, rapid innovation in the private sector and a commercial research and development sector dwarfing that of the U.S. military create new challenges for the U.S., which will need to quickly identify and integrate cutting edge technological developments in this rapidly changing environment.

Because of the foundational impact autonomy and artificial intelligence will have on the character of warfare, CNA created the Center for Autonomy and Artificial Intelligence to focus on these emerging technologies and their contribution to national security. The Center capitalizes on the ability to leverage the scientists and analysts of CNA's staff of nearly 700, with their experience base in military operations, test and evaluation, security and intelligence analyses, technology assessment, and autonomy and AI.

ABOUT CNA CORPORATION

CNA is a not-for-profit research and analysis organization with 75 years of experience providing government agencies with data-driven insights and real-world, actionable solutions grounded in our direct experience with the operational environments where these solutions are applied. CNA developed the foundational techniques for operational analysis to address complex challenges facing government programs. We have applied these techniques successfully in areas ranging from defense to aviation, education, justice, and homeland security.

For more information please contact:
Dr. Larry Lewis, Director, Center for Autonomy and AI
703-824-2020
Lewisl@cna.org