Artificial Intelligence and Global Security Initiative
Research Agenda
July 2017

The Center for a New American Security's Artificial Intelligence and Global Security Initiative explores how the artificial intelligence (AI) revolution could lead to changes in global power, the character of conflict, and crisis stability. The Initiative will also examine the security dimensions of AI safety and prospects for international cooperation.

Key research questions include:

Shifting Power Dynamics

- How might AI alter power dynamics among relevant actors in the international arena? (great and rising powers, developed countries, developing countries, corporations, international organizations, militant groups, other non-state actors, decentralized networks and movements, individuals, and others)
- How will geopolitical, bureaucratic, cultural, or other factors affect how actors choose to adopt AI technology for military or security purposes?
- Will some types of power (e.g., military, economic, informational, political) become more significant than others in a world with extensive AI systems?
- How might the key components of power shift during an AI revolution? What resources could become more or less valuable (e.g., large datasets on which to train learning algorithms)?
- How rapidly should we expect AI technology to proliferate among states and non-state actors? Are some forms of AI subject to more rapid proliferation than others? How will this influence the first mover advantages and fast follower potential from developing AI technologies?
- To what extent are AI technologies inherently dual use? What will influence the relative applicability of AI to the military and commercial sectors?
- How can "progress" in AI research effectively be tracked and measured? What progress points would signal important technological milestones or the need for a change in approach?

The Character of Conflict

- How might military applications of AI change the character of warfare?
- Does AI empower different actors? How will that affect which state and non-state actors end up in conflicts?

- How will actors deploying AI systems fight? Does AI change the means and methods of conflict? How might AI affect factors such as speed, controllability, transparency, lethality, or physical and psychological distance in war?
- Will AI change the relative utility of different resources in ways that will change the aims of conflict?
- How might the use of AI change military structures, recruiting, and retention? How might these changes further exacerbate other economic shifts taking place within society? Will some people increasingly be left behind, losing opportunity for advancement?
- What investments in R&D could affect military and intelligence AI applications?
- Do changes in the character of conflict challenge and/or require changes in the ethical and legal frameworks for conflict, military or otherwise?
- Is there a point at which AI-enabled systems might change the nature of war itself?

Crisis Stability

- How might AI-led changes to the character of conflict affect conflict initiation and escalation, including war outbreak, termination, and escalation control?
- Are AI-enabled technologies likely to alter: (1) first-strike stability, potentially incentivizing states to strike first in a crisis; (2) defense against attack, making aggression more or less costly; or (3) escalation control, making it more or less difficult for policymakers to limit conflicts?
- How will the proliferation of AI systems influence the probability of arms races, and how might those arms races in particular influence crisis stability?
- Are there especially dangerous or risky forms of AI or applications of AI that might undermine crisis stability? What factors might influence whether states or other actors adopt these approaches?

AI Safety

- How do concerns about AI safety/control apply to national security applications of AI? Do security applications of AI (e.g., malware, cyber defenses, military robotics, etc.) minimize or exacerbate concerns about AI safety?
- What are the implications of the vulnerability of AI-enabled systems to manipulation by competitors (e.g., "fooling" images/data, data poisoning, behavioral hacking)? To what extent are AI-enabled systems vulnerable to subversion?
- How are various actors who apply AI for security purposes likely to approach AI safety? Are some actors more incentivized than others to take safety concerns seriously?

- How likely is a "race to the bottom" in safety for security applications of AI? How might various actors react to how others approach AI safety? To what extent does AI safety require cooperation?
- What regulatory and other government approaches can prevent AI technologies from being misused?
- How might geopolitical, cultural, or other factors affect different actor's desire for safety and accuracy with AI technologies?

Prospects for Cooperation

- What incentives are there for cooperation and/or competition among various actors? What do these incentives depend on?
- How does the current state of openness among the AI research community affect prospects for cooperation or competition? How would a change in openness affect incentives among various actors?
- How might various actors react to sudden developments or shocks – accidents, technology surprise, or sudden shifts in power dynamics? Can these shocks be anticipated or planned for?
- Which potentially harmful consequences of AI on global security require cooperation to be avoided and which can be avoided through unilateral action by various actors?
- Are there steps that can be taken now that might make cooperation to avoid harmful outcomes more or less likely?

More information on the Artificial Intelligence and Global Security Initiative can be found at cnas.org/ai.