AI Governance in Military Affairs

AI Ethical Principles: Implementing US Military's Framework

By Megan Lamberth

SYNOPSIS

The US Department of Defence (DoD) has made real progress in recent months on the development and adoption of ethical principles in artificial intelligence (AI). The challenge now is establishing a framework for how AI ethics can be institutionalised across the DoD's myriad components and missions.

COMMENTARY

THE US DEFENCE Department is still in the early stages of determining how best to ensure the development and deployment of AI that is ethical, reliable, and secure. Earlier this year, the DoD formally adopted a set of AI ethical principles that are meant to guide the Department's development, adoption, and use of AI-enabled systems.

The DoD, and in particular, the Joint Artificial Intelligence Centre (JAIC), is in the midst of transforming those principles into actionable guidance for DoD personnel. For the principles to be meaningful and enduring, the JAIC will need additional authority and resources; the DoD will also need to work hand-in-hand with allies and partners who are also tackling the challenge of ensuring safe, secure, and ethical AI. What can others learn from the US experience?

Adopting and Implementing AI Principles

Over the last two years, the DoD has taken a number of steps to lay the groundwork for AI adoption. In June 2018, the DoD established the JAIC to serve as the synchronising body for AI activities across the entire Department, including the military

services. Eight months later, the DoD's first AI strategy called for the development of "resilient, robust, reliable, and secure" AI systems and vowed to articulate a set of principles to ensure AI capabilities were used in a "lawful and ethical manner".

The task of formulating those ethical principles fell to the Defence Innovation Board (DIB), a federal advisory committee of tech experts and private industry leaders that counsel senior Defence Department leaders on emerging technology. The DIB released their proposed AI principles in October 2019, and three months later, the DOD formally adopted them. In short, the Defence Department vows to develop and deploy AI systems that are:

1. *Responsible*. Responsibility for the "development, deployment, and use" of AI-enabled systems rests with the human operator.

2. *Equitable*. The DoD will work to "minimise unintended bias" in its AI capabilities.

3. *Traceable*. DoD personnel involved in the lifecycle of an AI system will have an "appropriate understanding" of the technology, its development, and its operational use.

4. *Reliable*. The DoD will ensure its AI systems have "explicit, well-defined uses" and that these systems will undergo testing and evaluation through its entire operational lifecycle.

5. *Governable*. AI systems will have the capability to "detect and avoid unintended consequences" and "disengage or deactivate deployed systems that demonstrate unintended behavior."

Signal to US Allies

The adoption of these AI principles was a notable success for the Defence Department. The principles serve as a signal to private tech companies and US allies that the DoD cares not only about the adoption of new AI capabilities, but *how* those capabilities are used. The principles are broad and to a certain extent vague, but this was done intentionally.

AI is an enabling technology with potential use cases across warfighting, command and control, logistics, maintenance, and more. The principles had to be broad enough to encompass future AI use, as well as AI applications that are already in widespread use throughout the Defence Department. As former director of the JAIC, Lt. General Shanahan explained: "Tech advances, tech evolves…the last thing we wanted to do was put handcuffs on the department to say what we could and could not do."

The JAIC's job now is to transform these principles into actionable guidance that is malleable enough to address a variety of applications, but concrete enough to facilitate meaningful action across the Department. The JAIC has launched two new initiatives that are working to develop frameworks for ethical AI and a "shared vocabulary" for all DoD personnel working on AI.

The Responsible AI Subcommittee — part of the DoD's broader working group on AI

— is an interdisciplinary group composed of DoD representatives working to establish the necessary ethical frameworks for the acquisition and implementation of AI. The [Responsible AI Champions pilot](#), a programme designed to train DoD personnel on the importance of ethics in AI, is exploring how the Department's ethical principles can be actualised across the lifecycle of an AI system.

Way Forward

Despite these important strides, the JAIC still faces headwinds in its attempt to implement the Department's guidance on AI ethics. The JAIC's foremost challenge is reconciling the imbalance of its extensive mandate with its current level of authority and resources. As a [RAND report](#) from December 2019 found, the JAIC "does not have directive or budget authorities, and that critically limits its ability to synchronise and coordinate DoD-wide AI activities to enact change".

To fulfill its mission, the JAIC needs additional resources — both budgetary and personnel — and greater authority. While the JAIC's staff has steadily increased since its founding in 2018, it must continue growing in order to successfully manage the Department's integration of AI. The JAIC also needs its own acquisition authority to hasten the process of [acquiring and testing](#) new AI capabilities.

A [House version](#) of the draft fiscal year 2021 National Defence Authorisation Act (NDAA) calls for this change in authorities. The same [NDAA draft](#) would move the JAIC up in the chain-of-command by placing it directly under the deputy secretary of defence, a necessary change that would give the JAIC increased visibility across the Department.

Bringing in Like-Minded Countries

In addition to increasing the JAIC's authorities and resources, the DoD should continue seeking buy-in and participation from allied and partner countries, many of whom are wrestling with similar challenges of incorporating ethics into AI development. The JAIC's [partnership](#) with Singapore's Defence Science and Technology Agency (DSTA) is illustrative of the benefits of collaboration on AI development.

Last year, the two countries held a multi-day series of exchanges centered on the use of AI for [disaster and humanitarian relief](#). Earlier this year, the JAIC announced the first ["AI Partnership for Defence"](#) — an initiative aimed at bringing [like-minded countries](#) together to share best practices on incorporating ethics into AI systems.

These kinds of partnerships are critical, particularly as the geostrategic competition between the US and China continues to evolve. Close collaborations, like the partnership between the JAIC and DSTA or the AI Partnership for Defence, allow for greater [interoperability](#) between militaries, access to global [datasets](#), and the sharing of best practices to ensure the development and fielding of responsible AI. Ultimately, these partnerships will benefit the security of not only the US and allied countries in the Indo-Pacific region, but countries around the globe.

Challenges Ahead

The obstacles obstructing the process of implementation will not be resolved overnight. The Defence Department will continue to reckon with how to ensure the security and reliability of its AI systems as it continues to develop and field new capabilities. At the same time, as my colleague Martijn Rasser and I wrote earlier this year, the DoD will have to overcome the institutional and bureaucratic barriers that stifle its ability to adopt and widely deploy AI capabilities.

The JAIC's mission for now is to make sure the Department's guidance on ethics and AI is clear and widely understood, so that DoD personnel from logistics to warfighting to intelligence know not only what ethical AI is, but how it should be applied.

Megan Lamberth is a research assistant with the Technology and National Security Program at the Center for a New American Security (CNAS). This is part of an RSIS Series in association with the Military Transformations Programme.