WORKSHOP BRIEF

BEYOND ARMS CONTROL: REGULATING DEFENSE AND SECURITY AI TECHNOLOGIES

Artificial intelligence (AI) technologies are growing more prevalent in security and defense applications, particularly among national governments, law enforcement organizations, for-profit companies, and private citizens. However, the power and pervasiveness of AI technologies threaten to disrupt relations between nations, as well as the broader international community. Discussions to date about AI defense technologies have focused on the treatment of lethal autonomous weapons systems (LAWS). Experts have argued that the range of conflict-related applications of AI beyond LAWS requires careful consideration and a multifaceted approach.

On July 15-16, 2019, CIFAR, in collaboration with the leadership team (below), convened 20 participants from government, academia, and non-profits to scope new approaches to governing AI-powered systems in conflict situations.

IMPACTED STAKEHOLDERS

- Humanitarian organizations and civil society
- Technologists and engineers
- State security and military organizations
- Non-state and paramilitary actors
- Private technology corporations
- Academics and researchers

KEY INSIGHTS

1. The application of AI in the context of weaponry is unique in that AI is not itself a weapon, but rather a technology that can be incorporated into other systems. Not only does this complicate traditional approaches to arms control, but it also involves unconventional actors, such as a wider range of private firms beyond arms manufacturers.
2. The use and governance of defense and security technologies have traditionally been monopolized by nation states. The increasing availability of AI has reduced the barriers to access, which makes these technologies much more accessible to actors beyond state security organizations.
3. AI's reliance on data allows for AI-powered systems to operate without human interaction or interference. However, human biases and oversights can influence how these machines operate as a result of the way optimization protocols are designed, or through the use of biased or inaccurately labeled data.
4. Export control is not reliable as many autonomous systems have already spread worldwide, and the technology itself is highly accessible and applicable by a larger group of actors.
5. The breadth of AI applications cannot be addressed solely through traditional international security bodies and mechanisms, such as the UN Group of Governmental Experts and Convention on Certain Conventional Weapons.
6. Cooperative security agreements have arms control components that can be applied to autonomous weapons systems, as they are preventative rather than reactive in nature. These agreements, such as the Cooperative Threat Reduction Directive, rely on collaborative efforts between nations, emphasizing information sharing, joint activities, and coordinating efforts.

RECOMMENDATIONS AND NEXT STEPS

1. Guidelines and standards of best practice ought to be multilaterally and multisectorally developed to contribute to meaningful human control of AI-powered weapons and engage a more diverse group of stakeholders.
2. Keeping humans in the loop is necessary to oversee machine operations and can help identify biases or design oversights in data and AI systems. Ensuring meaningful human oversight can reduce the risks associated with unsupervised machines acting on biased data outcomes.
3. Standards and certification can help to prevent abusive and reckless use by placing certification controls at each step of production training.
4. It is necessary to engage and inform industry developers in order to have them identify the potential risks of proposed systems, and to ensure they adhere to both domestic and international standards. Moreover, incentives need to be created for businesses to adhere to standards of responsible AI practices and deployment.
5. Efforts need to be made to address the disconnect between technologists and policymakers. Technical literacy is low among policymakers, while technologists and engineers often do not understand public policy, law, or international governance; addressing these knowledge gaps and creating venues for collaboration are important for engaging all stakeholders in a meaningful way.

LEADERSHIP TEAM

Kerstin Vignard, UN Institute for Disarmament Research (UNIDIR), Switzerland; David Danks, Carnegie Mellon University, United States; Amb (ret.) Paul Meyer, Simon Fraser University, Canada

FURTHER READING

- Regulating Autonomous Systems: Beyond Standards
- Defense Research and Development Canada Military Ethics Assessment Framework
- United Nations Convention on Certain Conventional Weapons

CIFAR

RÉSUMÉ DE L'ATELIER

AU-DELÀ DU CONTRÔLE DES ARMES : RÉGLEMENTATION DES TECHNOLOGIES DE DÉFENSE ET DE SÉCURITÉ BASÉES SUR L'IA

Les technologies basées sur l'intelligence artificielle (IA) sont de plus en plus répandues dans les applications de sécurité et de défense, en particulier au sein des gouvernements nationaux, des forces de l'ordre, des entreprises à but lucratif et chez les particuliers. Cependant, la puissance et l'omniprésence des technologies basées sur l'IA menacent de perturber les relations entre les pays et entre les membres de la communauté internationale. Jusqu'à présent, les discussions sur les technologies utilisées dans le domaine de la défense se sont concentrées sur le traitement des systèmes d'armes létales autonomes (SALA). Les experts ont fait valoir que l'éventail des applications basées sur l'IA s'étendait au-delà des SALA et exigeait un examen attentif et une approche multidimensionnelle.

Les 15 et 16 juillet 2019, le CIFAR, en collaboration avec l'équipe de direction (ci-dessous), a convoqué 20 participants provenant des gouvernements, du milieu universitaire et d'organismes sans but lucratif afin de définir de nouvelles approches pour la gouvernance des systèmes basés sur l'IA dans des situations de conflit.

INTERVENANTS TOUCHÉS

- Gouvernements et responsables des politiques
- Organisations humanitaires et société civile
- Technologues et ingénieurs
- Organisations militaires et chargées de la sécurité de l'État
- Acteurs non étatiques et paramilitaires
- Sociétés technologiques privées
- Universitaires et chercheurs

FAITS SAILLANTS

1. L'application de l'IA dans le contexte de l'armement est unique en ce sens que l'IA n'est pas en soi une arme, mais plutôt une technologie qui peut être intégrée à d'autres systèmes. En plus de compliquer les approches traditionnelles en matière de contrôle des armements, cette situation implique des acteurs non conventionnels, comme un large éventail de firmes privées autres que les fabricants d'armes.
2. L'utilisation et la gouvernance des technologies de défense et de sécurité ont traditionnellement été monopolisées par les États-nations. La disponibilité croissante de l'IA a levé les obstacles, et ces technologies sont maintenant beaucoup plus accessibles aux acteurs autres que les organisations chargées de la sécurité de l'État.
3. L'utilisation de données permet aux systèmes basés sur l'IA de fonctionner sans interaction ni interférence humaine. Cependant, les biais et les erreurs des humains peuvent influencer le fonctionnement de ces machines en raison de la façon dont les protocoles d'optimisation sont conçus ou de l'utilisation de données biaisées ou identifiées de façon imprécise.
4. Le contrôle des exportations n'est pas fiable puisqu'un grand nombre de systèmes autonomes ont déjà été implantés dans le monde et que la technologie elle-même est très accessible et utilisable par un groupe plus nombreux d'acteurs.
5. En raison de leur variété, les applications basées sur l'IA ne peuvent être contrôlées uniquement par des organismes et des mécanismes de sécurité internationaux traditionnels, comme le Groupe d'experts gouvernementaux de l'ONU (en anglais) et la Convention sur certaines armes classiques de l'ONU.
6. Les accords de coopération en matière de sécurité comportent des composantes sur le contrôle des armements qui peuvent s'appliquer aux systèmes d'armes autonomes, car ils sont de nature préventive plutôt que réactive. Ces accords, comme la Directive sur la réduction concertée des menaces (en anglais), reposent sur des efforts de collaboration entre les pays, qui mettent l'accent sur le partage d'information, les activités conjointes et la coordination des efforts.

RECOMMANDATIONS ET PROCHAINES ÉTAPES

1. Des lignes directrices et des pratiques exemplaires devraient être mises au point de façon multilatérale et multisectorielle afin de favoriser le contrôle humain des armements fondés sur l'IA et de mobiliser un groupe plus diversifié de parties prenantes.
2. Il est impératif que les humains continuent de surveiller le fonctionnement des machines et aident à déceler les biais ou les erreurs dans la conception des systèmes de données et d'IA. Assurer une surveillance humaine peut réduire les risques associés aux machines non surveillées qui agissent à partir de données biaisées.
3. Les normes et la certification peuvent contribuer à prévenir les utilisations abusives et imprudentes en mettant en place des contrôles à chaque étape, de la formation à la production.
4. Il est nécessaire de mobiliser et d'informer les développeurs de l'industrie afin qu'ils cernent les risques potentiels des systèmes proposés et qu'ils se conforment aux normes nationales et internationales. De plus, des mesures incitatives doivent être instaurées pour encourager les entreprises à adhérer à des normes régissant des pratiques et un déploiement responsables en matière d'IA.
5. Des efforts doivent être consentis pour combler le fossé entre les technologues et les responsables de politiques. La littératie technique est faible chez ces derniers, tandis que les technologues et les ingénieurs ne saisissent pas toujours les politiques publiques, les lois ou la gouvernance internationale. Il est important de combler ces lacunes et de créer des occasions de collaboration afin de mobiliser toutes les parties prenantes de manière significative.

ÉQUIPE

Kerstin Vignard, Institut des Nations Unies pour la recherche sur le désarmement (UNIDIR) (Suisse) ; David Danks, Université Carnegie Mellon (États-Unis) ; Paul Meyer (ambassadeur [à la retraite]), Université Simon Fraser (Canada)

LECTURES COMPLÉMENTAIRES

- Regulating Autonomous Systems: Beyond Standards (en anglais)
- Defense Research and Development Canada Military Ethics Assessment Framework (en anglais)
- United Nations Convention on Certain Conventional Weapons (en anglais)