DROIT & POLITIQUES
HUMANITAIRES

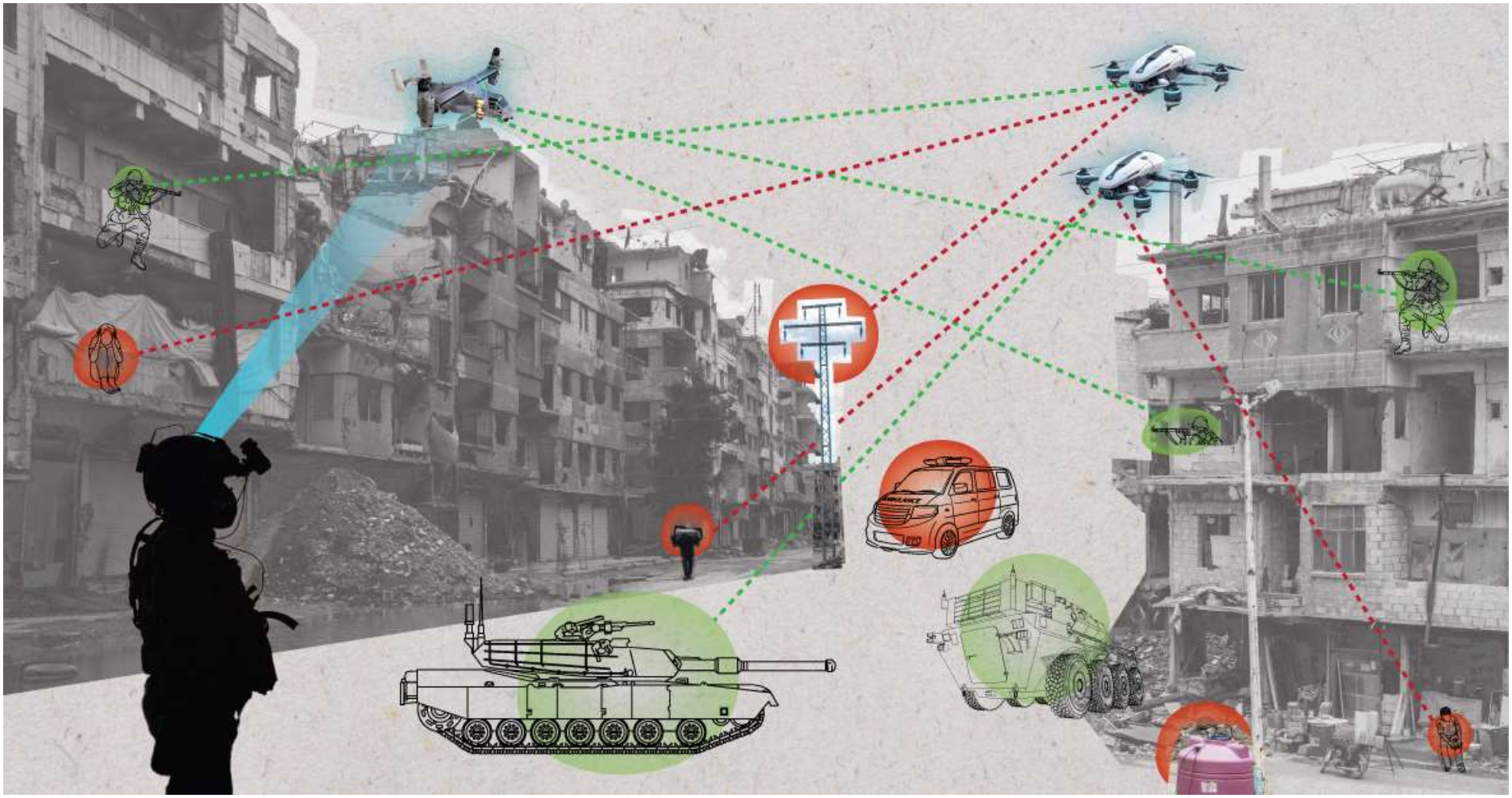
Les algorithmes de la guerre : le recours à l'intelligence artificielle pour la prise de décision dans les conflits armés

février 22, 2024, Action humanitaire / Nouvelles technologies

20 minutes de lecture

Ruben Stewart
Conseiller sur les forces armées et les groupes armés, CICR

Georgia Hinds
Conseillère juridique, CICR



En moins d'un an, Chat-GPT est devenu un nom familier, reflétant les incroyables progrès des logiciels dotés d'une intelligence artificielle (IA), en particulier les modèles d'IA générative. Ces progrès se sont souvent accompagnés de prédictions selon lesquelles l'IA allait révolutionner la guerre. À ce stade du développement de l'IA, nous en sommes encore au stade de découverte des modalités d'applications possibles, mais l'intérêt militaire pour la technologie de l'IA est indéniable. Dans son *livre blanc sur la défense nationale*, la Chine a prôné la théorie de l'« intelligentisation » de la

guerre, affirmant que l'exploitation de l'IA est un élément clé de la stratégie de modernisation de l'Armée populaire de libération. Le directeur de l'Agence américaine de cybersécurité et de sécurité des infrastructures (*Cybersecurity and Infrastructure Security Agency*, CISA), Jen Easterly, a souligné le fait que l'intelligence artificielle pourrait être « *l'arme la plus puissante de notre époque* ». Par ailleurs, si les *systèmes d'armes autonome*s ont régulièrement dominé les débats sur les usages militaires de l'IA, l'utilisation de cette dernière dans les systèmes d'aide à la prise de décision humaine dans les conflits armés est passée plus inaperçue.

Dans cet article, Ruben Stewart, conseiller du CICR sur les forces armées, et Georgia Hinds, conseillère juridique, font l'examen critique de certains des avantages prétendus de l'IA lorsqu'elle est utilisée pour éclairer la prise de décision des acteurs armés pendant la guerre. Les auteurs abordent notamment la question de la réduction des dommages causés au sein de la population civile et celle de la vitesse d'exécution opérationnelle, en accordant une attention particulière aux conséquences sur la population civile dans les conflits armés.

Même avant le récent engouement autour de l'IA, vous aviez probablement déjà utilisée une IA sous diverses formes et il se peut que vous lisiez cet article sur un appareil doté de cette technologie. Vous avez déverrouillé votre téléphone grâce à votre empreinte digitale ou à la reconnaissance faciale, interagi sur les réseaux sociaux, planifié des déplacements à l'aide d'une application téléphonique ou acheté quelque chose en ligne, qu'il s'agisse d'une pizza ou d'un livre : vous avez certainement eu recours à l'IA. À bien des égards, nous nous sommes habitués à l'IA, l'invitant, souvent à notre insu, dans notre quotidien.

Mais que se passerait-il si ce logiciel de reconnaissance faciale servait à identifier une personne à attaquer ? Et si, au lieu de trouver le vol le moins cher pour une destination donnée, un logiciel similaire sélectionnait un avion pour lancer une frappe aérienne sur une cible ? Ou si, au lieu d'indiquer la meilleure pizzeria ou le taxi le plus proche, la machine recommandait des plans d'attaque ? Il semblerait que cette réalité soit « bientôt disponible » selon les *entreprises qui développent des plateformes décisionnelles fondées sur l'IA dans un objectif de défense*.

Ces systèmes d'aide à la décision fondés sur l'IA sont des outils informatisés qui utilisent des logiciels d'IA pour afficher, synthétiser et/ou analyser des données et, dans certains cas, formuler des recommandations – voire des prédictions –– afin d'éclairer la prise de décision humaine en temps de guerre.

Les avantages des systèmes d'aide à la décision fondés sur l'IA sont souvent présentés en termes d'amélioration de la compréhension de la situation et d'accélération de la prise de décision. Ces arguments sont analysés ci-après, en tenant compte aussi bien des limites des systèmes fondés sur l'IA que des limites humaines et dans le cadre du processus de planification des conflits contemporains.

Réduire le risque de porter atteinte aux populations civiles dans les conflits

L'avènement de nouvelles technologies de guerre s'accompagne souvent de l'argument selon lequel leur intégration réduira les dommages causés au sein de la population civile (bien que cela *ne se vérifie pas toujours dans la pratique*). Il a par exemple été affirmé que les systèmes d'aide à la décision fondés sur l'IA pourraient contribuer à mieux *protéger les civils dans les conflits, dans certaines circonstances*. Certes, le *droit international humanitaire (DIH)* oblige les commandants militaires et les autres personnes responsables d'exécuter des opérations à fonder leurs décisions sur l'évaluation des informations de toute origine dont ils disposent sur le moment. Lors de combats en zone urbaine en particulier, le CICR a *recommandé* que les informations sur des éléments tels que la présence de civils et de biens de caractère civil proviennent également de sources d'information ouvertes telles qu'internet. Par ailleurs, concernant en particulier l'IA et l'apprentissage automatique, le CICR a conclu que, dans la mesure où les *outils d'aide à la prise de décision fondés sur l'IA peuvent permettre de collecter et d'analyser ce type d'informations plus rapidement et à plus grande échelle*, ils pourraient tout à fait permettre *aux êtres humains de prendre de meilleures décisions dans les conflits et de réduire au minimum les risques pour les civils*.

En même temps, chaque diagnostic produit par un système d'aide à la décision fondé sur l'IA *doit faire l'objet de vérifications croisées pour se prémunir contre les informations biaisées ou inexactes*. Ce principe, qui s'applique à toute source d'information recueillie en période de conflit, est encore plus important lorsque la prise de décision s'appuie sur l'IA ; comme le CICR *l'a déjà souligné*, il peut être extrêmement difficile – voire impossible – de vérifier la pertinence de ces diagnostics en raison du fonctionnement du système et de la manière dont s'organise l'interaction homme-machine. Ces aspects sont examinés plus en détail ci-après.

Les limites du système

Ces derniers temps, les médias ont souvent mentionné l'évolution de l'IA en faisant état de ses échecs, parfois catastrophiques : par exemple, des logiciels *qui ne reconnaissent pas ou identifient mal les personnes à la peau plus foncée* ou *qui recommandent des itinéraires de voyage sans tenir compte d'informations routières actualisées*, ou encore des cas d'accidents mortels provoqués par des *véhicules* autonomes. Certaines de ces défaillances peuvent s'expliquer, sans que cela ne les justifie, par le fait que les données sur lesquelles les systèmes se fondent sont *biaisées*, corrompues, *empoisonnées* ou tout simplement fausses. Ces systèmes *peuvent également être facilement « trompés » ; des techniques peuvent être utilisées pour induire le système à mal classifier les données*. Par exemple, dans un conflit, l'ennemi pourrait très bien employer des techniques pour *modifier le code source d'un système d'aide au ciblage afin qu'il identifie les bus scolaires comme des véhicules ennemis*, ce qui aurait des conséquences dévastatrices.

Lorsque l'IA est utilisée pour des tâches plus complexes, en particulier lorsque l'analyse, mais aussi potentiellement la prise de décision et l'évaluation, sont faites à différents niveaux simultanément, il est presque impossible de vérifier le diagnostic et l'origine de toute erreur potentielle ayant contribué à ce résultat final. À mesure que la complexité des systèmes s'accroit, le risque de cumuler les erreurs augmente – *une petite erreur dans la première recommandation algorithmique se répercute sur un deuxième processus algorithmique qu'elle vient fausser, puis sur un troisième, et ainsi de suite*.

C'est pourquoi les systèmes fondés sur l'IA se comportent souvent d'une manière qui ne peut être expliquée ni par l'utilisateur ni par le développeur, même après une analyse postérieure plus approfondie. *Une étude portant sur le très médiatisé modèle linguistique GPT-4* a révélé que sa capacité à résoudre un problème mathématique avait drastiquement et inexplicablement diminué, passant de 83,6% à seulement 35,2% à trois mois d'intervalle. *Les comportements imprévisibles*

peuvent également découler de l'apprentissage par renforcement, les machines s'étant révélées très efficaces pour adopter et dissimuler des comportements imprévus, et parfois négatifs, afin de déjouer l'esprit humain et le surpasser, que ce soit en mentant pour être en position de force dans la négociation ou en *prenant des raccourcis pour battre un jeu sur ordinateur*.

Les défis de l'interaction homme-machine

Les systèmes d'aide à la décision fondés sur l'IA ne « prennent » pas de décisions. En revanche, ils influencent directement et souvent de manière significative les décisions des êtres humains, notamment en raison des limites et des tendances humaines sur le plan cognitif en cas d'interaction homme-machine.

Ainsi, le « biais d'automatisation » désigne la tendance humaine à ne pas faire d'examen critique des diagnostics générés par un système ou à s'abstenir de rechercher des informations contraires – *en particulier dans les situations d'urgence*. Ce phénomène a déjà été observé dans d'autres domaines tels que la santé, où les *résultats erronés de l'IA ont nui à la précision des diagnostics posés par des radiologues expérimentés*.

Des diagnostics inexacts dans le domaine de la santé peuvent avoir une issue fatale. De même, dans un conflit armé, l'excès de confiance peut entraîner des conséquences funestes. En 2003, le système de défense américain *Patriot* a tiré à deux reprises sur des avions appartenant à des membres de la coalition, parce qu'ils avaient été désignés à tort comme des missiles de l'ennemi. Une enquête réalisée après les faits avait identifié comme l'une des principales insuffisances le fait que « *les opérateurs étaient formés à faire confiance au logiciel du système* ».

Ces modes de fonctionnement, conjugués aux caractéristiques de l'interaction homme-machine, augmentent potentiellement la probabilité que les diagnostics générés par des IA s'écartent de l'intention du décideur humain. En temps de guerre, cela peut entraîner une *escalade accidentelle* et, en tout état de cause, accroître les *risques pour les civils et les personnes protégées*.

La maîtrise de la vitesse d'exécution

L'un des avantages militaires vantés de l'IA est qu'elle permet d'avoir une boucle décisionnelle plus rapide que l'adversaire. Une plus grande vitesse d'exécution engendre souvent des risques supplémentaires pour les civils, c'est pourquoi des techniques pour allonger ces délais, à l'instar de la « patience stratégique », sont employées pour réduire les pertes en vies humaines. Ralentir la prise de décision, y compris les processus et les évaluations visant à l'éclairer, offre au système et à l'utilisateur davantage de temps pour :

- avoir une vision élargie de la situation
- mieux appréhender la situation, et
- envisager plus d'options.

On notera que cela se vérifie tout au long de la chaîne décisionnelle et pas seulement au « point de prise de décision » final. Par conséquent, l'argument selon lequel l'aide à la prise de décision fondée sur l'IA permet en réalité de laisser *davantage* de temps pour la patience stratégique, en accélérant les étapes chronophages menant à la décision finale de « presser ou non la détente », comporte le risque de simplifier à l'excès le processus de ciblage et l'usage de la force dans les conflits contemporains.

Avoir plus de temps permet d'avoir une vision élargie

La désormais tristement célèbre attaque de drone lancée sur Kaboul le 29 août 2021 pendant l'évacuation de la ville et qui a tué 10 civils, a été attribuée par le commandant du Commandement central au fait de ne pas avoir « eu suffisamment de temps pour étudier les modes de vie et prendre un certain nombre d'autres mesures [traduction CICR] ».

L'analyse des « modes de vie » une expression utilisée par certaines forces armées pour désigner ce qui consiste à évaluer la présence et le nombre de civils et de combattants, leurs habitudes de vie, leurs déplacements quotidiens, *etc.*, dans une zone que l'on prévoit d'attaquer et dans ses environs. C'est *un moyen efficace pour réduire les dommages causés au sein de la population civile*. Cela étant, évaluer les modes de vie ne peut se faire qu'en temps réel, c'est-à-dire dans le temps nécessaire aux civils pour mettre en pratique ces modes de vie et le processus ne peut être accéléré.

Les tentatives de prédire des comportements futurs en se fondant sur les habitudes passées des populations ne permettront pas de prendre en compte la situation actuelle. Dans cet exemple, l'examen de documents des renseignements plus anciens, notamment des vidéos de Kaboul, n'aurait pas révélé l'évolution de la situation et des comportements après la prise de pouvoir des Talibans et les mesures d'évacuation qui étaient en cours.

Comme le soulignent *les directives sur la prévention des pertes civiles*, « plus vous attendez et prenez le temps d'observer la situation, plus vous serez au courant de ce qui se passe et préparés à prendre la décision d'employer des moyens létaux ou non létaux [traduction CICR] » ou, pour paraphraser Napoléon, « habillez-moi lentement, je suis pressé » – on obtient parfois les meilleurs résultats en restant placide.

Avoir plus de temps permet de mieux appréhender la situation

Une autre raison de ralentir la vitesse d'exécution de la prise de décision est que l'être humain a besoin de temps, d'une part pour bien comprendre une situation, en particulier lorsque celle-ci est chaotique et difficile, d'autre part pour choisir la réponse appropriée à apporter. Lorsqu'il ne dispose pas du temps nécessaire, l'être humain est moins capable d'appréhender correctement la situation. Le processus de planification militaire est conçu pour donner aux commandants et au personnel militaire le temps d'étudier le contexte opérationnel, l'adversaire, les forces alliées et la population civile, ainsi que les avantages et les inconvénients des plans d'action envisagés. La compréhension acquise à travers ce processus de réflexion ne peut être externalisée car, comme l'a affirmé le général Dwight D. Eisenhower, « chaque fois que je me prépare au combat, je me dis que les plans sont inutiles, mais la planification fait tout [traduction CICR] ».

Cela a des conséquences lorsque les décideurs humains doivent se prononcer sur un plan d'action généré ou « recommandé » par un système d'aide à la prise de décision fondé sur l'IA. Sa capacité à accélérer la vitesse d'exécution opérationnelle par rapport à celle de l'ennemi est probablement la raison la plus invoquée

pour justifier son emploi. Sans avoir mis en œuvre le plan d'action proposé par un tel système, ni même l'avoir totalement compris, le planificateur humain aura sans doute une connaissance limitée de la situation, des différents facteurs d'influence et des acteurs impliqués. En effet, on a constaté que *le recours à une assistance automatisée peut* réduire la vigilance des utilisateurs humains et nuire à leur capacité à rester attentif dans telle ou telle situation. Cette question doit être examinée au regard de son influence sur le respect des obligations prévues par le DIH ; de l'*obligation de faire tout ce qui est pratiquement possible pour vérifier que les cibles* sont licites découle l'obligation de s'appuyer autant que possible sur des services de renseignement et sur les moyens de surveillance et de reconnaissance disponibles afin d'acquérir la meilleure connaissance possible de la situation dans ces circonstances.

Avoir plus de temps permet d'envisager plus d'options

Outre le fait qu'il donne au commandant une vision élargie et une meilleure compréhension de la situation, ce délai supplémentaire favorise la mise en place d'*alternatives stratégiques*, pouvant conduire à décider de ne pas faire usage de la force ou d'apaiser les tensions. Avoir plus de temps permet aux autres unités et plateformes de se désengager, se repositionner, se réapprovisionner, planifier une opération à venir et se préparer à y participer. Le commandant dispose ainsi de davantage d'options, y compris de solutions alternatives susceptibles de mieux réduire les atteintes portées à la population civile. Ce délai supplémentaire peut permettre de prendre d'autres mesures d'atténuation, telles que la diffusion d'alertes. Il permet aussi à la population civile de mettre en œuvre des mécanismes d'adaptation, par exemple en se mettant à l'abri, en se réapprovisionnant en nourriture et en eau ou en évacuant la zone.

Prenons l'exemple de la doctrine sur la planification militaire des forces armées américaines qui énonce que « si l'on dispose de délais suffisants et qu'il n'y a aucun avantage à agir plus rapidement, rien ne justifie de ne pas prendre le temps de procéder à une planification adéquate [traduction CICR] ». En effet, comme le rappelle le manuel de l'OTAN relatif à la protection des civils, « *lorsque l'on dispose du temps nécessaire pour bien planifier les actions, pour respecter la discrimination et prendre précisément pour cible une force ou un objectif conformément aux principes du DIH, les risques de [pertes civiles] sont fortement réduits* ».

Conclusion

« *La guerre est chaotique, meurtrière et elle est par essence le propre de l'homme. C'est un affrontement entre différentes volontés qui se joue au sein d'une population et entre ses membres. Au fond, toute guerre vise à changer les comportements humains, chaque camp essayant de modifier le comportement de l'autre par la force des armes* [traduction CICR] ». Les guerres résultent de désaccords humains, opposent des groupes d'êtres humains, sont contrôlées par des êtres humains et ce sont ces êtres humains qui, au lendemain des conflits, doivent coexister. Plus important encore, ce sont les êtres humains qui endurent les souffrances causées par les hostilités.

Cette réalité, de même que le DIH, appelle une *approche « centrée sur l'humain »* à l'égard du développement et de l'emploi de l'IA dans les conflits armés – pour tenter de préserver une certaine humanité, au sein de qui est déjà inhumain. Cette approche comporte au moins deux aspects essentiels : 1) l'accent mis sur les personnes qui pourraient être affectées ; et 2) l'accent mis sur les obligations et les responsabilités des personnes qui font appel ou ordonnent de faire appel à l'IA.

S'agissant des personnes qui pourraient être affectées, il n'importe pas seulement d'atténuer les risques auxquels est confrontée la population civile lorsque l'on recourt à des systèmes d'aide à la prise de décision fondés sur l'IA pour obtenir un avantage militaire. Il existe également la possibilité de concevoir et d'utiliser ces outils spécialement dans le but de protéger les civils. Parmi les pistes possibles, figurent des outils permettant de reconnaître et de suivre la présence de populations civiles et d'en alerter les forces armées, ou de reconnaître des emblèmes distinctifs indiquant un statut protégé dans un conflit armé (voir *ici* et *ici*).

En outre, pour que les êtres humains puissent s'acquitter des obligations qui leur incombent en vertu du DIH, les systèmes d'aide à la prise de décision peuvent être utiles pour éclairer la décision humaine, mais ne sauraient la remplacer dans des situations qui mettent en péril la vie et la dignité des personnes dans les conflits armés. Une grande partie des États l'admettent concernant les systèmes d'armes autonomes (voir, par exemple, *ici*, *ici* et *ici*). C'est aux individus et à ceux qui les commandent qu'incombe la responsabilité de respecter le DIH et non aux ordinateurs. Selon le manuel du droit de la guerre du département américain de la Défense, « *Le droit de la guerre ne prévoit pas l'obligation que les armes prennent des décisions juridiques [...]. Ce sont les personnes qui doivent se conformer au droit de la guerre [traduction CICR] »*. La Chine a souligné cette question de manière plus générale dans ses *normes éthiques pour l'intelligence artificielle de nouvelle génération*, en insistant sur le fait que « les humains sont ceux qui sont responsables en dernier ressort ».

Les arguments selon lesquels les systèmes d'aide à la prise de décision fondés sur l'IA permettront nécessairement d'améliorer la protection des civils et le respect du DIH doivent faire l'objet d'un examen critique et être analysés en prenant en compte ces considérations, compte tenu de ce que nous savons des limites de ces systèmes, de l'interaction homme-machine et des conséquences de l'accélération de la vitesse d'exécution des opérations.

Cet article a été initialement publié en anglais le 24 octobre 2023.

Voir aussi :

- Joelle Rizk, Sean Cordey, *Les menaces numériques dans les conflits armés : ce qui nous échappe et comment y remédier*, 2 octobre 2023.
- Laura Bruun, *Les systèmes d'armes autonomes : ce que le droit dit – et ne dit pas- sur le rôle de l'être humain dans l'usage de la force*, 18 janvier 2022

Tags: AI, armed conflict, artificial intelligence, ChatGPT, civilians, cyber defence, cyber security, decision-making, military, military cyber operations, new technologies

Ceci pourrait vous intéresser

Au-delà de l'accès : trois observations relatives à la sécurité alimentaire et à la prévention de la famine en période de conflit armé

17 minutes de lecture

Action humanitaire / Nouvelles technologies

Ariana Lopes Morey, Menty Kebede & Matt Pollard

Ces dernières années, des millions de personnes vivant dans des régions affectées par un conflit ont été confrontées à une insécurité alimentaire grave ...

COVID-19, conflits armés et violences sexuelles : renverser la charge de la preuve

7 minutes de lecture

Action humanitaire / Nouvelles technologies Sophie Sutrich

Bien que les violences sexuelles dans les conflits armés soient interdites par le droit international humanitaire, elles demeurent une cruelle réalité. À l'occasion de la ...