

Trouvez le Sala ! (1)

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

Éric GERMAIN

Membre de la Cerna (2). Il s'exprime ici à titre personnel et ses propos n'engagent ni la Direction générale des relations internationales et de la stratégie (DGRIS) pour laquelle il travaille, ni le ministère des Armées.

La campagne internationale contre les « robots tueurs » commence à véritablement intéresser les médias français (3) après la publication fin juillet 2015 d'une lettre ouverte. Elle est signée par d'éminentes personnalités du monde scientifique, au premier rang desquelles l'astrophysicien Stephen Hawking, et plusieurs patrons-stars de la Silicon Valley.

Si beaucoup de nos concitoyens découvrent alors ce nouveau sujet d'anxiété collective, un débat était déjà engagé depuis plusieurs années. En France, les Armées sont le premier acteur public à s'interroger sur les enjeux sociétaux nés des progrès du numérique et de la robotique et, sur le plan international, le ministère de la Défense – aujourd'hui ministère des Armées – a joué un rôle pionnier en lançant l'ouverture d'un nouveau chapitre du dialogue international de maîtrise des armements.

Cette généalogie du débat éthique, juridique et sociétal soulevé par la perspective de futurs Systèmes d'armes létaux autonomes (Sala) mérite d'être rappelée. Elle permet de mieux comprendre le travail mené ces cinq dernières années au sein de la Convention pour certaines armes classiques (CCAC) hébergée par l'Office des Nations unies à Genève. Elle éclaire surtout la stratégie de quelques industriels philanthropes qui opèrent avec cette campagne humanitaire un « jeu de bonneteau » où la focalisation autour des applications militaires de l'Intelligence artificielle (IA) permet opportunément de fixer l'angoisse des opinions publiques.

Nous proposons ici un essai d'anatomie d'une mobilisation internationale suivie, de l'intérieur, depuis près de dix ans.

(1) En référence au jeu de dupes du bonneteau, dont le nom en anglais est *Find the Lady*, « Trouvez la Dame », car il se joue traditionnellement avec trois cartes : les rois de pique et de trèfle, et la reine de cœur.

(2) Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene (l'Alliance qui regroupe l'INRIA, le CNRS, la CPU, la CDEFI, le CEA, etc.).

(3) En quelques semaines, cette pétition a suscité près de 200 articles dans la seule presse française et plus de 2 500 dans les médias anglophones selon un recensement effectué par Didier Danet au Centre de recherches des Écoles de Saint-Cyr Coëtquidan (CREC Saint-Cyr).

2009-2012 : l'ICRAC, une prise de conscience encore limitée au monde universitaire

2009 représente une année clef dans la prise de conscience internationale des enjeux inédits posés par les usages militaires des nouvelles technologies de la robotique et du numérique. En France, la « robotique autonome à usage militaire » est l'un des sujets qui motive la proposition de création d'un comité d'éthique en Sciences et Technologies du numérique ⁽⁴⁾. Du côté américain, trois ouvrages parus en 2008-2009 éveillent les consciences : *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, le best-seller de Peter W. SINGER ; *Governing Lethal Behavior in Autonomous Robots* de Ronald C. ARKIN ; et, sur un éventail plus large, *Moral Machines: Teaching Robots Right from Wrong* coécrit par Wendell WALLACH et Colin ALLEN ⁽⁵⁾.

Si cette prise de conscience s'exprime surtout aux États-Unis, c'est que ce pays est le premier (avec Israël ⁽⁶⁾) à avoir massivement investi dans les technologies dites *unmanned* (sans homme à bord) et cela, dans des proportions assez spectaculaires. Comme le rappelait Peter W. Singer (p. 32), aucun système robotisé n'est déployé au moment de l'entrée des troupes américaines en Irak en 2003, mais ils sont 150 dès l'année suivante pour atteindre, en 2008, une capacité de plus de 7 000 drones aériens et plus de 12 000 robots terrestres.

La plupart de ces systèmes ne sont alors pas armés et, quand ils le sont, l'ouverture du feu reste du ressort d'un opérateur humain qui agit à distance. Cependant, en mai 2009, le document de planification des évolutions capacitaires de la technologie des drones publié par l'Armée de l'air américaine affiche parmi ses dix perspectives majeures celle de passer d'un contrôle direct du système, *man in the loop*, à un contrôle supervisé, *man on the loop* ⁽⁷⁾. L'*US Air Force* annonce que, d'ici quelques décennies, les drones aériens seront capables de cibler et d'ouvrir le feu sur des combattants sans que cette action soit décidée par un opérateur humain. Par ailleurs, le document évoque des capacités futures de vitesse (hypersonique), de furtivité ou de saturation (essais de drones) propres à exciter les imaginations. Le rapport prospectif de l'*US Air Force* suscite des inquiétudes d'autant plus vives qu'il est publié le jour même où le directeur de la *CIA*, Leon Panetta, s'exprime pour vanter le rôle des drones armés dans la guerre contre le terrorisme ⁽⁸⁾. Pour la première fois, la *CIA* assume pleinement cette guerre clandestine menée avec une technologie robotisée en affirmant que les frappes de drones sont la seule solution efficace, « *the only game in town* », pour combattre *Al-Qaïda* au Pakistan et ailleurs.

(4) *Rapport sur la création d'un comité d'éthique en Sciences et Technologies du Numérique*, INRIA, mai 2009, 31 pages (www.cnrs.fr/comets/IMG/pdf/03-inriastn.pdf), cf. p. 15, point 2.3.2. Ce rapport aboutira à la création de la Cerna par Allistene.

(5) Respectivement publiés par Penguin (2009, 400 pages), Chapman and Hall/CRC (2009, 256 pages) et Oxford University Press (2008, 288 pages).

(6) Lors de la guerre du Kippour de 1973, les forces armées israéliennes sont les premières à utiliser des drones aériens dans des missions offensives épuisant le feu des batteries antiaériennes égyptiennes ; KENT Allen et WILLIAMS James G. (dir.), *Encyclopedia of Computer Science and Technology*, vol. 29, n° 14, 1993, p. 310.

(7) *United States Air Force Unmanned Aircraft Systems Flight Plan 2009-2047*, Headquarters, US Air Force, Washington, 18 mai 2009, 82 pages (www.govexec.com/pdfs/072309kp1.pdf), p. 14.

(8) BENSON Pam, « US airstrikes in Pakistan called 'very effective' », *CNN*, 18 avril 2009 (<http://edition.cnn.com/2009/POLITICS/05/18/cia.pakistan.airstrikes/>).

Cette approche décomplexée et validée par le président Obama ⁽⁹⁾ entraîne, dès l'année suivante, un usage clandestin de drones armés d'une ampleur absolument inédite. Cette « guerre des drones » est soutenue par une grande majorité de l'électorat américain, démocrate comme républicain ⁽¹⁰⁾. Même si certaines ONG qualifient ces systèmes d'armes de « drones tueurs », l'opinion publique semble rassurée par un usage télé-opéré maintenant toujours un être humain pour déclencher le tir, dans une chaîne de commandement validée au plus haut niveau de l'État ⁽¹¹⁾. Le discours rassurant l'opinion publique sur la présence pérenne de l'« homme dans la boucle » est cependant mis à mal au début de l'été 2010 par l'annonce du déploiement au bord de la zone démilitarisée entre les deux Corée d'un robot garde-frontière *SGR-A1* capable d'ouvrir le feu de manière automatisée sur tout individu repéré par son système de « tracking » ⁽¹²⁾. Pour certains, l'apparition de ce « robot sentinelle » ouvrant le feu sans supervision humaine démontre que la problématique n'est plus l'évolution lointaine d'un *man in the loop* à un *man on the loop*, mais que nous serions déjà à l'étape du *man out of the loop* ! La perspective d'un « homme en dehors de la boucle décisionnelle » d'une arme létale ne relevant plus du débat prospectif, il devenait donc urgent de prendre des mesures immédiates pour contrôler, voire enrayer, le développement de ces « robots tueurs ».

En cette « année-zéro des guerres robotisées » ⁽¹³⁾, l'*International Committee for Robot Arms Control (ICRAC)* organise à Berlin, du 20 au 22 septembre 2010, une réunion internationale devant réfléchir à la mise en place d'un « contrôle des armements pour les robots, télé-opérés et autonomes » ⁽¹⁴⁾. Plusieurs grands pays participent en envoyant des observateurs travaillant au sein des institutions de défense – l'éthicien américain Edward Barrett ⁽¹⁵⁾ ou moi-même – ou des experts de *think tanks* liés aux administrations, à l'exemple du *Royal United Services Institute (RUSI)* de Londres pour Elizabeth Quintana ou encore du Centre pour le contrôle des armements de Moscou dans le cas d'Eugene Miasnikov. Les principales voix du débat public outre-Atlantique sont également présentes (Ronald C. Arkin, Wendell Wallach, Colin Allen). On note aussi la participation d'acteurs religieux comme l'Institut pour la religion et la paix de l'aumônerie militaire autrichienne ⁽¹⁶⁾ ou encore le *Rowntree Charitable Trust*, une

(9) Dans un discours prononcé le 25 mars 2010, le Conseiller juridique du département d'État Harold KOH justifie la politique d'exécutions ciblées par drones et étend aux activités clandestines d'une agence civile de renseignement le concept de *Pre-Emptive Self Defense* de la *Global War on Terror* lancée par George W. Bush.

(10) MICAH Zenko, « U.S. Public Opinion on Drone Strikes », *Council on Foreign Relations*, 18 mars 2013 (www.cfr.org/blog/us-public-opinion-drone-strikes).

(11) Barack Obama s'est appuyé sur le *Patriot Act* pour conduire cette politique d'« exécutions ciblées » d'individus dont la liste lui est soumise lors d'une réunion hebdomadaire (*Terror Tuesdays*) ; BECKER Jo et SHANE Scott, « Secret 'Kill List' Proves a Test of Obama's Principles and Will », *New York Times*, 29 mai 2012 (www.nytimes.com/).

(12) RABIROFF Jon, « Machine gun-toting robots deployed on DMZ », *Stars and Stripes*, 12 juillet 2010 (www.strips.com/news/pacific/korea/machine-gun-toting-robots-deployed-on-dmz-1.110809).

(13) GERMAIN Éric, « 2010 : année zéro des guerres robotisées », *Revue Défense Nationale*, n° 740, mai 2011, p. 119-121 et, en anglais, « Outlook for Change in the International Legal Framework for the Use of Military Robotics », in DANET Didier, *Robots on the Battlefield*, Combat Studies Institute Press, Fort Leavenworth, 2013, p. 67-71 (www.law.upenn.edu/live/files/3882-hagerott-cyber-robots-history-and-warpdf).

(14) « Arms Control for Robots—Limiting Armed Tele-Operated and Autonomous Systems ».

(15) Edward T. Barrett, directeur de recherche au sein du *Stockdale Center for Ethical Leadership* de l'École navale d'Annapolis avait organisé cette même année (22-23 avril 2010) une conférence sur le thème « New Warriors & New Weapons: The Ethical Ramifications of Emerging Technologies ».

(16) Mandaté par l'Église catholique en 2009, l'*Institut für Religion und Frieden* a conduit une vaste étude sur les enjeux juridiques et éthiques des armes robotisées dirigée par Gerard Dabringer présent à Berlin.

fondation du mouvement Quaker dont l'aide financière permet à l'*ICRAC* d'organiser cette manifestation ⁽¹⁷⁾.

Si l'inquiétude face à l'accélération des développements technologiques est partagée par l'ensemble des participants, en revanche la réponse à y apporter fait l'objet d'un vrai débat. Faut-il se mobiliser pour demander un moratoire sur les armes « autonomes » ? Celui-ci doit-il porter sur la recherche ou seulement sur les développements ? Comment prendre en compte le caractère éminemment dual des technologies en cause, le cas du robot sentinelle sud-coréen développé par une société connue pour ses télévisions et téléphones portables – Samsung – n'illustre-t-il pas cette difficulté ? Enfin, une interdiction est-elle pragmatiquement envisageable ?

Avec l'intervention de son conseiller médical, le Dr Mark Steinbeck, la division juridique du Comité international de la Croix Rouge (CICR) montre aux participants toute la complexité d'un régime de contrôle pour des armes robotisées dont les règles d'engagement et les effets attendus sont encore très incertains. L'*ICRAC*, fondé un an auparavant par quatre chercheurs ⁽¹⁸⁾, oriente les débats vers deux options ⁽¹⁹⁾ : le moratoire ou l'interdiction totale, la seconde réunissant finalement la majorité des suffrages.

Outre le lancement formel de la campagne « *Arms Control for Robots* », une matinée de discussion avec des membres du *Bundestag* a sans doute contribué à renforcer l'intérêt des partis de gauche allemands pour ce sujet. Cela aura un impact important à partir de la fin de 2013 lorsque les sociaux-démocrates (*SPD*) et Les Verts mettront l'engagement pour l'interdiction des armements autonomes au cœur des accords de « grande coalition » des cabinets Merkel III et Merkel IV.

Néanmoins, sur le moment, l'impact médiatique de cet événement est objectivement très faible tout comme l'intérêt que cette campagne internationale suscite dans ses deux premières années d'existence. Les grandes ONG spécialisées dans les questions de désarmement suivent cette initiative, mais sans s'y impliquer ⁽²⁰⁾. Dans un premier temps, la campagne de l'*ICRAC* n'a aucune influence sur la communication des armées américaines qui, pour la première fois en 2011, affiche des modèles de drones terrestres et aériens armés sur la couverture de sa *Feuille de route* bisannuelle consacrée à la robotique militaire ⁽²¹⁾.

(17) Dans la deuxième phase du mouvement, on trouvera aussi un rôle important de *PAX*, la branche néerlandaise océanique de l'ONG *Pax Christi* tout comme, au sein de la CCAC, l'action discrète mais influente de Mission du Saint-Siège aux Nations unies.

(18) L'*ICRAC* est créée au Royaume-Uni en septembre 2009 par le roboticien britannique Noel Sharkey, le physicien allemand Jürgen Altmann et deux philosophes américain et australien Peter Asaro et Robert Sparrow.

(19) Le professeur Ronald C. ARKIN propose une troisième voie dans l'article collectif écrit dans les semaines suivant la réunion de Berlin. Nous revenons en conclusion sur cet article paru en 2011 dans la *Columbia Science and Technology Law Review*.

(20) Joanne Mariner, la directrice du programme « Terrorisme et Contre-terrorisme » au sein de *Human Rights Watch*, a participé à la réunion de Berlin de septembre 2010.

(21) *Unmanned Systems Roadmap FY2011-2036*, document publié depuis 2007 par le ministère de la Défense américain (<https://info.publicintelligence.net/DoD-UAS-2011-2036.pdf>). Le choix de mettre des matériels robotisés armés en couverture est repris pour l'édition suivante de 2013 (<https://info.publicintelligence.net/DoD-UnmannedRoadmap-2013.pdf>).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

Le faible impact initial de la campagne de l'ICRAC peut être mis sur le compte d'au moins deux facteurs. D'une part, elle rencontre des problématiques internes liées à la capacité financière très modeste de l'ICRAC et au manque de professionnalisme de ses membres qui présentent plutôt un profil de chercheurs et d'universitaires, plus enclins à utiliser la raison que l'émotion dans leur argumentaire. Or, une campagne médiatique de désarmement en appelle d'abord à l'émotion de l'opinion publique (amplifier le sentiment de dégoût, de réprobation morale spontanée, ce que la Prix Nobel de la paix 1997 Jody Williams qualifie de « *yuck* effect » ou « effet "beurk" »). Pour y arriver, il faut souvent simplifier la problématique et user d'une vérité sélective, voire approximative, deux critères compliqués à mettre en œuvre pour un esprit scientifique, même pour servir une cause juste. La seconde raison de la faible mobilisation internationale des premières années tient aussi à un « positionnement » trop futuriste, appelant à une mobilisation contre des armes qui n'existent pas encore. À cela s'ajoute un monde militant pacifiste qui, dans ces années, est surtout mobilisé par l'espoir que suscitent les discours du président Barack Obama annonçant l'engagement de l'Amérique pour « un monde sans armes nucléaires » ⁽²²⁾.

Cependant, l'actualité scientifique va relancer l'intérêt médiatique puis associatif. En février 2011, le programme informatique Watson remporte le jeu Jeopardy et certains de ses concepteurs chez IBM évoquent la réalisation du test de Turing (la capacité de conversation en langage naturel entre l'homme et la machine), seuil qui permettrait d'entrevoir la création d'une « IA forte ». La perspective angoissante d'une machine dotée d'une forme de conscience d'elle-même capable de surpasser l'homme relance l'intérêt médiatique. Il permet à Noel Sharkey d'entamer à la fin de l'année 2011 le rapprochement de l'ICRAC avec plusieurs grandes ONG, dont *Human Rights Watch* (HRW). Tout au long de l'année 2012, il réussit à convaincre des militants aguerris des processus d'Ottawa et d'Oslo (pour l'interdiction des mines antipersonnel et des armes à sous-munitions) de rejoindre cette nouvelle cause contre les « robots tueurs ».

2013-2015 : HRW, Christof Heyns et l'introduction du débat par la France à la CCAC

La seconde étape de professionnalisation de la campagne *Stop Killer Robots* se prépare tout au long de 2012 avec le soutien apporté en mars par l'initiative britannique *Article 36* ⁽²³⁾ et son fondateur Richard Moyes qui organise le plan stratégique d'un projet de coalition d'ONG scellé en octobre à New York sous le pilotage de HRW. Un mois plus tard, HRW publie un rapport intitulé *Losing Humanity—the case against Killer Robots*, un document d'une cinquantaine de pages dont le roboticien Noel

(22) Le 4 avril 2009, le président Obama annonce à Prague : « *America's commitment to seek the peace and security of a world without nuclear weapons* » alimentant une dynamique qui se poursuit jusqu'à son message du 27 avril 2015, lors de la Conférence de révision du Traité de non-prolifération nucléaire (TNP), où il rappelle la phrase prononcée à Berlin en 2013 : « *so long as nuclear weapon exist we are not truly safe* ».

(23) BOLTON Matthew, NASH Thomas et MOYES Richard, « Ban autonomous armed robots », 5 mars 2012 (www.article36.org/statements/ban-autonomous-armed-robots/). Fondée quelques mois auparavant par Richard Moyes et Thomas Nash, le nom de cette ONG fait explicitement référence à l'article n° 36 intitulé « Armes nouvelles » du 1^{er} Protocole additionnel (PA I) aux Conventions de Genève (1949) signé en 1977.

Sharkey, le fondateur de l'ICRAC est le principal conseiller technique. HRW reprend la même stratégie adoptée vingt ans auparavant lorsqu'elle s'était appuyée sur l'expertise de l'ONG *Physicians for Human Rights* pour rédiger le rapport *Landmines in Cambodia—The Coward's War* de 1991 qui précéda le lancement de campagne internationale contre les mines antipersonnel.

Alors que le rapport *Losing Humanity* est mis en ligne le 19 novembre, par une heureuse coïncidence le Pentagone publie, à peine 48 heures après, sa toute première doctrine sur l'autonomie des systèmes d'armes ⁽²⁴⁾. L'un des officiers auteurs de ce texte, Paul Scharre, rejoint dès l'année suivante un *think tank* de Washington, le *Center for a New American Security* (CNAS) pour continuer à travailler sur cette thématique.

On peut imaginer que le département de la Défense américain avait quelque idée de la réflexion menée au sein des bureaux new-yorkais de HRW, comme de celle conduite aux Nations unies par Christof Heyns, le Rapporteur spécial sur les exécutions extrajudiciaires. Ce dernier présente son rapport au Conseil des droits de l'homme de l'ONU le 9 avril 2013 ⁽²⁵⁾. Il s'inquiète de la perspective de développement et de prolifération d'armes nouvelles appelées LARs pour « *Lethal autonomous robotics/robots* », reprenant à son compte une inquiétude déjà exprimée par son prédécesseur (Philip Alston) dans son rapport de 2010. Mais le juriste sud-africain va plus loin en « engageant les États à imposer des moratoriums nationaux sur certaines activités relatives au LARs ».

HRW convie à Londres les 22 et 23 avril 2013 une trentaine d'ONG susceptibles d'être intéressées par ce nouveau sujet de maîtrise des armements. Elle souhaite les convaincre de l'intérêt du sujet et bâtir une stratégie de campagne. Les médias et les représentants institutionnels ne sont pas conviés mais, à l'invitation de l'ICRAC et du Mouvement Pugwash, nous y avons assisté. Pour la seconde fois, la campagne contre les « robots tueurs » est lancée mais, cette fois-ci, avec une approche beaucoup plus stratégique.

Outre-Manche comme outre-Atlantique, il est courant de parler de *philantropy industry* pour qualifier l'activité d'ONG humanitaires qui affichent un professionnalisme assumé et même revendiqué tant dans la gestion des capitaux dont elles disposent que pour les méthodes de *marketing*, de *lobbying* ou de *social influencing* qu'elles peuvent adopter. L'ONG *Human Rights Watch* fait partie de cette catégorie et c'est sans abus de langage que l'on peut qualifier d'« OPA amicale » sa prise de contrôle de la campagne initialement conduite par l'ICRAC. L'arrivée de cette ONG répond au choix stratégique de son bureau new-yorkais car HRW avait raté la mobilisation sur les « drones tueurs » (peut-être en raison de sa proximité avec l'Administration Obama ⁽²⁶⁾) et se

(24) DEPARTMENT OF DEFENSE, *Directive n° 3000.09 « Autonomy in Weapon Systems »*, 21 novembre 2012 (<https://cryptome.org/dodi/dodd-3000-09.pdf>) ; ce document avait été précédé par un travail amont du DoD DEFENSE SCIENCE BOARD, *The Role of Autonomy in DoD Systems*, Task Force Report, juillet 2012.

(25) HEYNS Christof, Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, 09 avril 2013, United Nations General Assembly, 22 pages, (www.ohchr.org/) cf. p. 7, point n° 35.

(26) L'un des cadres de HRW, Tom Malinowski, est nommé en avril 2014 au poste de Secrétaire d'État-adjoint pour la démocratie et les droits de l'Homme.

trouvait distancée parmi les ONG du désarmement. En s'investissant dans un sujet de désarmement « futuriste », *HRW* inversait cette erreur de positionnement en ringardisant à son tour la mobilisation contre les drones présentés comme « la “Ford modèle T” des nouvelles technologies de guerre »⁽²⁷⁾.

Le fait de parler d'un système d'armes qui n'existe pas et pour lequel on ne peut donc pas présenter d'image de « robot tueur », et encore moins de ses victimes, semblait freiner l'intérêt médiatique nécessaire pour créer une mobilisation populaire. Pourtant, *HRW* réussit le tour de force de transformer ce handicap en atout. Comme le déclare Steve Goose, le directeur de la division Armes de *HRW*, devant les ONG réunies à Londres, dans ce nouveau type de campagne « sans survivants, sans victimes et sans visuels », il est inutile de dépenser de l'énergie et de l'argent pour présenter les dangers de cette technologie, l'expression de « robots tueurs » renvoyant chacun à un imaginaire anxigène, nourri par des décennies de cinéma hollywoodien.

Sauver l'humanité d'un futur à la *Terminator* par l'adoption d'un *Pre-Emptive Ban* (une « interdiction préventive ») est une proposition difficilement critiquable. Pour autant aiguillonne-t-elle la mobilisation publique vers les applications les plus immédiatement préoccupantes de ces technologies ? On pouvait répondre par la négative en 2013⁽²⁸⁾ et les cinq années suivantes nous ont plutôt confortées dans ce sentiment. Pour autant, le sujet de futurs systèmes d'armes « autonomes » dans leurs fonctions de ciblage et d'ouverture de feu présente-t-il des enjeux éthiques essentiels que les militaires ont le devoir de se poser ? La réponse est incontestablement positive.

La perspective de disposer de robots sans émotions (sans peur, sans haine ou sans fatigue) qui, programmés pour respecter les Conventions de Genève, seraient déployés en lieu et place de soldats nous semble susciter davantage l'intérêt de civils que de militaires. On aurait tort d'y voir une simple réaction « corporatiste » de soldats craignant que leur métier ne soit un jour confié à des machines. Ils comprennent mieux que d'autres qu'un être humain qui accepte sous l'uniforme de son pays de donner la mort, mais aussi le risque de la recevoir, n'a rien à voir d'un point de vue éthique avec une « machine » qui (à ce jour) ne fait pas un choix, mais calcule un choix. Or, une décision potentiellement létale reposant sur un calcul statistique peut être vue comme portant atteinte à la dignité humaine.

Avant même que la campagne *Stop Killer Robots* ne soit relancée à Londres, une sénatrice française avait posé une question écrite au ministre de la Défense pour connaître la doctrine française en matière d'armes totalement autonomes également surnommées « robots tueurs »⁽²⁹⁾. Dans sa réponse publiée le 25 juillet 2013, Jean-Yves Le Drian affirme que « les enjeux éthiques, juridiques et sociétaux que pose la robotique dans ses usages militaires, mais également civils, sont discutés très en amont

(27) Interview de Jody WILLIAMS à *Time* le 15 mars 2013 diffusée sur *YouTube* (www.youtube.com/), cf. 3'35.

(28) GERMAIN Éric, « La campagne contre les “robots tueurs” se trompe-t-elle de cible ? », *Le Monde*, 24 juin 2013 (www.lemonde.fr/) et aussi GUIBERT Nathalie, « Le spectre des robots tueurs », *Le Monde*, 18 juin 2013 (www.lemonde.fr/).

(29) GARRIAUD-MAYLAM Joëlle (Sénatrice UMP des Français établis hors de France), Question écrite n° 4498 au ministre de la Défense sur les armes totalement autonomes soumise le 07 février 2013, réponse publiée le 25 juillet 2013 (www.nossenateurs.fr/question/14/04498).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

et de manière critique et responsable au sein du ministère de la Défense ». Il ajoute que « si ces questions venaient à être portées dans une enceinte multilatérale, la France privilégierait le cadre onusien de la Convention sur certaines armes classiques, adoptée à Genève le 10 octobre 1980 par la Conférence des Nations unies, qui vise à réduire les souffrances humaines en établissant un cadre normatif permettant de mieux contrôler l'utilisation de certains matériels spécifiques ».

Le ministre indique ici clairement son souhait de voir le sujet porté dans une enceinte universelle de discussion. Après quelques échanges avec le ministère des Affaires étrangères, il est effectivement apparu que seule une approche multilatérale permet d'envisager une réponse effective à des problématiques éthiques et sociétales globales et que la Convention de 1980 sur certaines armes classiques (CCAC ou *CCW* en anglais) représente le cadre le plus légitime pour cela.

La France, prenant la présidence annuelle de la CCAC à l'automne 2013, c'est donc assez naturellement que notre représentant permanent auprès de la Conférence du désarmement à Genève organise le 3 septembre un premier séminaire sur les enjeux de futures armes autonomes. En novembre 2013, lors de l'assemblée générale annuelle de la CCAC, il propose à ses homologues de mettre ce nouveau sujet à l'ordre du jour. La première étape est d'ouvrir un dialogue informel, ouvert aux experts, afin de permettre aux États parties au traité de se faire une idée du périmètre du sujet et de ses enjeux. Il est décidé de ne pas parler de « robots létaux autonomes » mais de « systèmes d'armes létaux autonomes » (Sala ou *LAWS* en anglais) afin de ne pas stigmatiser une discipline technologique particulière (la robotique), d'autant plus que le sujet clef de l'autonomie relève davantage du domaine du numérique que de la robotique. Fin 2013, les *LARs* deviennent des *LAWS* !

Le mandat adopté en novembre 2013 précise que « sous l'autorité du président, une réunion informelle d'experts de quatre jours se tiendra du 13 au 16 mai 2014 afin de discuter des questions relatives aux technologies émergentes liées aux systèmes d'armes létaux autonomes ». En lançant le sujet à la CCAC, la France partage le point de vue du CICR qui, dans le rapport de sa 31^e Conférence internationale, précise que le passage d'un système d'armes « automatisé » à un système d'armes « autonome » signifie deux choses : « une capacité d'apprentissage et d'adaptation » et une « intelligence artificielle »⁽³⁰⁾.

La réunion d'experts de 2014 permet de commencer un débat ouvert à une grande variété de points de vue. L'objectif initial est de comprendre ce que pourrait être un Sala, un exercice prospectif délicat dans la mesure où il faut analyser les progrès technologiques les plus récents pour extrapoler une ou plusieurs évolutions possibles. Il s'agit bien évidemment d'une réflexion assez technique, dont la complexité ne satisfait ni les ONG, ni les États soucieux de générer une couverture médiatique importante.

(30) ICRC, *Report of the 31st International Conference of the Red Cross and Red Crescent*, octobre 2011, p. 39 (www.icrc.org/). Une définition reprise par LAWAND Kathleen, « Fully autonomous weapon systems », 25 novembre 2013 (www.icrc.org/eng/resources/documents/statement/2013/09-03-autonomous-weapons.htm).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

Dans les mois qui suivent cette première réunion d'experts à la CCAC, le ministère des Affaires étrangères allemand finance ⁽³¹⁾ des travaux de recherche susceptibles d'encadrer les débats de la prochaine présidence allemande de la CCAC. Mais cette recherche d'expertise se fait majoritairement auprès de militants de la campagne pilotée par *HRW* qui souhaitent abandonner la distinction autonomie/automatisme pour ne parler que de « niveaux d'autonomie » dont l'« autonomie totale » sera le degré ultime ⁽³²⁾. L'ambassadeur d'Allemagne qui succède à la présidence française reprend à son compte le concept de « contrôle humain significatif » que l'ONG Article 36 avait proposé dès le lancement de la campagne contre les « robots tueurs » ⁽³³⁾. Cette notion de *Meaningful Human Control (MHC)* appliquée aux « fonctions critiques » de ciblage et d'ouverture du feu des armements permet de sortir du cadre du mandat initial (celui des systèmes autonomes). Le *MHC* peut concerner tout type d'armement automatisé car l'appréciation d'un contrôle « significatif » est suffisamment subjective pour concerner *a priori* n'importe quel système d'armes moderne.

Sans surprise, les deux réunions d'experts sur les Sala de 2014 et de 2015 sont marquées par un raidissement des positions entre des pays investis de longue date dans les campagnes de désarmement et les grandes puissances militaires très hostiles à s'engager dans des mesures contraignantes fondées sur une appréciation subjective. Pour une majorité de pays, un Sala demeure encore une abstraction qui reste à définir ou, du moins à caractériser (dire, notamment, ce que ne serait pas un Sala).

La couverture médiatique de ces débats prend aussi une orientation moins technique et plus politique. En 2010, le choix de tout organisateur d'une conférence qui aurait voulu confronter un « pro » et un « anti » Sala se serait naturellement porté sur deux roboticiens : Noel Sharkey et Ronald C. Arkin. À partir de 2014, il ne s'agit plus de roboticiens, mais d'intellectuels davantage rodés aux arcanes de la communication stratégique avec, par exemple, l'universitaire américaine Heather Roff face à Paul Scharre, ancien officier du Pentagone désormais responsable de l'*Ethical Autonomy Project* ⁽³⁴⁾ au sein du *CNAS*, un *think tank* de Washington.

De l'été 2015 à l'été 2018 : le tandem Stuart Russel/Elon Musk et la création du groupe gouvernemental d'experts à la CCAC

La troisième époque de la campagne internationale contre les Sala est marquée par l'entrée en scène d'un éminent professeur en intelligence artificielle, Stuart Russel,

(31) Un financement direct par une subvention à l'Institut des Nations unies pour la recherche sur le désarmement (*UNIDIR*) puis, à partir de 2017, par une aide indirecte à l'*IPRAWN (International Panel on the Regulation of Autonomous Weapons)* mis en place par la *Stiftung Wissenschaft und Politik (SWP)*.

(32) À l'exemple de la réflexion menée dans le domaine de l'automobile par la *Society of Automotive Engineers* proposant six degrés d'autonomies ; REESE Hope, « Autonomous driving levels 0 to 5 », *TechRepublic*, 20 janvier 2016 (www.techrepublic.com/article/autonomous-driving-levels-0-to-5-understanding-the-differences/).

(33) ARTICLE 36, « Killer Robots: UK Government Policy on Fully Autonomous Weapons », avril 2013 (www.article36.org/wp-content/uploads/2013/04/Policy_Paper1.pdf).

(34) Un projet rendu possible grâce à une généreuse donation de la *John D. and Catherine T. MacArthur Foundation* qui a financé le programme *20YY Warfare Initiative* dirigé par Paul Scharre. Au même moment, le *CNAS (Center for a New American Security)* accueille Robert Work qui sera le « père » de la *Third Offset Strategy* annoncée fin 2014.

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

d'un industriel visionnaire, Elon Musk, et par la médiatisation apportée par le parrainage du physicien Stephen Hawking.

Le *Future of Life Institute (FLI)* voit le jour en mars 2014, mais c'est en janvier 2015 que cette association prend une dimension internationale avec l'organisation à Puerto Rico de la conférence « L'avenir de l'IA : opportunités et défis ». Elle réunit du 2 au 5 janvier 2015 des scientifiques de renom, mais aussi des grands patrons de l'industrie du numérique dont Luke Nosek (PayPal), Elon Musk (SpaceX et Tesla Motors), Demis Hassabis (Google Deepmind) et Jaan Tallinn (Skype). Plusieurs spécialistes en communication stratégique sont invités comme Charina Choi, alors chargée de définir la « Vision et stratégie IA » de Google, ou Steve Crossan, un ancien cadre de l'ONG *Amnesty International* ayant rejoint Google en 2005 pour y monter des projets philanthropiques d'« impact social »⁽³⁵⁾.

Il n'y a pas eu de couverture médiatique de cette réunion, car « afin de faciliter des discussions candides et constructives, aucun média n'était présent »⁽³⁶⁾. Le sujet des Sala y a été vraisemblablement abordé par l'éthicien Wendell Wallach et surtout par la politologue Heather Roff Perkins, très impliquée la campagne *Stop Killer Robots*. Dans une *interview* donnée à son université de retour de cette conférence, elle déclare : « Je suis intervenue pour dire : voici le type d'armes que nous avons actuellement et voici ce vers quoi (le département de la Défense) dit qu'il se dirige pour les 20 ou 30 prochaines années »⁽³⁷⁾. Le 15 janvier 2015, Elon Musk annonce un don de 10 millions de dollars au *FLI* qui permet d'envisager une véritable action d'influence à l'échelle mondiale et d'orienter stratégiquement ses activités vers l'objectif de « développer des visions optimistes du futur »⁽³⁸⁾. Pour atteindre cet objectif de « faire aimer l'IA à ceux qui en ont peur », il s'agit désormais de fixer une peur générique sur un seul objet : l'IA militaire et son incarnation objectivement la plus anxiogène, le Sala.

Les universitaires qui alimentent la réflexion intellectuelle de la campagne *Stop Killer Robots* de *HRW* bénéficient directement de cette manne financière offerte par Elon Musk – 116 974 \$ pour Peter Asaro (*New School*, New York), 136 918 \$ pour Heather Roff (Université de Denver) ou les 180 000 \$ attribués à Wendell Wallach (Université de Yale)⁽³⁹⁾ –, mais ce sont surtout les chercheurs qui fournissent des réponses technologiques aux problématiques éthiques qui se trouvent le plus généreusement financés : 275 000 \$ pour Francesca Rossi (Université de Padoue) ou 342 727 \$ pour Stuart Russell (Université de Californie à Berkeley).

(35) Dans cette politique de « social impact », Steve Crossan est à l'origine en 2011 du soutien très médiatisé aux internautes égyptiens (l'initiative *speaktweet*) et de l'ouverture de l'Institut culturel Google de Paris.

(36) « AI safety conference in Puerto Rico », *Future of Life Institute*, 12 octobre 2015 (<https://futureoflife.org/2015/10/12/ai-safety-conference-in-puerto-rico/>).

(37) Propos recueillis par FORGUES Brent, « Korbel Professor Attends Innovative Conference on the Future of Artificial Intelligence », 21 janvier 2015 (www.du.edu/korbel/about/news/2015-ai-conference.html).

(38) *Mission: To catalyze and support research and initiatives for safeguarding life and developing optimistic visions of the future, including positive ways for humanity to steer its own course considering new technologies and challenges* (<https://futureoflife.org/team/>).

(39) TEGMARK Max, « Elon Musk donates \$10M to keep AI beneficial », *FLI*, 12 octobre 2015 (<https://futureoflife.org/2015/10/12/elon-musk-donates-10m-to-keep-ai-beneficial/>).

Présent à Puerto Rico, le professeur Russell participe à la seconde réunion informelle d'experts sur les Sala qui se tient à Genève du 13 au 17 avril 2015. Nous nous sommes rencontrés pour la première fois à cette occasion et il nous a expliqué (et répété à la tribune de la CCAC⁽⁴⁰⁾) que son intérêt pour la campagne anti-« robots tueurs » est motivé par le souci de préserver la réputation de sa discipline, l'intelligence artificielle. Le directeur du Centre pour les systèmes intelligents de l'Université de Californie à Berkeley regrette que désormais, à chaque fois qu'il aborde la question de l'IA, de l'apprentissage-machine ou de la robotique « intelligente », cela se termine inmanquablement par une discussion sur *Terminator*. Plus trivialement, et il ne s'en cache nullement, il craint que cette image délétère n'entraîne une baisse des financements des laboratoires de recherche travaillant dans cette discipline.

Stuart Russell en conclut que, pour redonner de la popularité aux recherches et aux applications civiles de l'IA, il convient de les distinguer clairement de leurs perspectives les plus anxiogènes qu'il identifie comme exclusivement liées aux domaines de la sécurité et de la défense. Dans cette action stratégique de « communication de crise », Stuart Russell trouve un appui majeur auprès des grands patrons de la Silicon Valley.

Le 28 juillet 2015, à l'ouverture de l'*International Joint Conference on Artificial Intelligence (IJCAI)*, le forum mondial de l'IA organisé depuis 1969) est lancée une pétition appelant à l'interdiction des développements militaires de l'IA qui conduiraient à la fabrication d'armes autonomes. Cette lettre ouverte reprend une première initiative lancée en janvier⁽⁴¹⁾ mais en se focalisant sur le seul domaine militaire avec des « armes autonomes [qui] deviendront les *Kalashnikov* de demain »⁽⁴²⁾. Le dernier paragraphe de cette seconde lettre ouverte synthétise l'approche manichéenne adoptée « *Good AI vs. Bad AI* » (distinguant la « bonne » IA civile de la « mauvaise » IA militaire) : « En résumé, nous croyons que l'IA présente un grand potentiel bénéfique pour l'humanité et c'est l'objectif que nous devons avoir dans ce domaine. Commencer une course aux armements dans l'IA militaire est une mauvaise idée et devrait être empêchée par une interdiction des armes offensives autonomes au-delà d'un contrôle humain significatif »⁽⁴³⁾.

On ne parle pas ici de Sala mais d'« armes de l'IA » (*AI weapons*) dépassant le seuil d'un « contrôle humain significatif ». Ce seuil subjectif doit être défini par des normes et c'est bien dans ce champ normatif que le *Future of Life Institute* entend désormais peser. Il va s'appuyer sur une association professionnelle ayant développé une grande force normative : *The Institute of Electrical and Electronics Engineers (IEEE)*

(40) Cf. diapo n° 24 de sa présentation Power Point « Artificial Intelligence: Implications for Autonomous Weapons » (www.unog.ch/).

(41) MACMILLAN Robert, « AI Has Arrived, and That Really Worries the World's Brightest Minds », *WIRED*, 16 janvier 2015 (www.wired.com/2015/01/ai-arrived-really-worries-worlds-brightest-minds/) et pour le texte de la première lettre ouverte (<https://futureoflife.org/ai-open-letter/?cn-reloaded=1>).

(42) Pour une critique détaillée de cette lettre ouverte et de ses assertions, lire GANASCIA Jean-Gabriel, POWERS Thomas M. et TESSIER Catherine, « On the Autonomy and Threat of "Killer Robots" », *APA Newsletter on Philosophy and Computers*, vol. 17, n° 2, printemps 2018, p. 3-9.

(43) AI & ROBOTICS RESEARCHERS, « Autonomous Weapons: An Open Letter » (<https://futureoflife.org/open-letter-autonomous-weapons/>).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

qui revendique plus de 400 000 membres répartis sur tous les continents. On retrouve, parmi les signataires de la lettre ouverte de juillet 2015, de nombreux membres d'*IEEE-CS*, la *Computer Society*, la plus importante des 39 sociétés techniques que compte *IEEE*. L'implication de l'association dans la campagne du *FLI* correspond à l'orientation stratégique donnée par son directeur de la branche standardisation et propriété intellectuelle qui souhaite « l'accroissement de l'influence de l'*IEEE* dans les domaines-clef techno-politiques, ce qui inclut de considérer les implications sociales et éthiques de la technologie »⁽⁴⁴⁾.

En mars 2016, la victoire du système AlphaGo de Google DeepMind contre le champion de Go sud-coréen Lee Sedol relance les inquiétudes sur la maîtrise humaine du progrès scientifique. Le 5 avril 2016, juste avant que ne s'ouvre la troisième réunion d'experts sur les Sala à Genève, *IEEE* lance sa *Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems*⁽⁴⁵⁾. Cette initiative mondiale forme un groupe de travail spécifique sur le sujet des « systèmes d'armes autonomes » présidé par le directeur du projet IA du *Future of Life Institute*⁽⁴⁶⁾. La quasi-totalité des membres du groupe de travail sont investis dans la campagne contre les armes autonomes : Noel Sharkey naturellement, mais aussi les trois chercheurs dont les travaux sont désormais financés par Elon Musk (Peter Asaro, Heather Roff et Stuart Russell) ainsi que Ryan Gariepi, le fondateur de Clearpath Robotics, une société canadienne qui est la première entreprise du secteur à rejoindre la campagne *Stop Killer Robots* en août 2014⁽⁴⁷⁾.

La *Global Initiative* de l'*IEEE* organise plusieurs évènements tout au long de l'année dont un « Forum éthique et IA » le 15 novembre 2016 à Bruxelles auquel nous avons assisté. Quelques semaines plus tard sort la première version d'un document intitulé : *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*⁽⁴⁸⁾. Les solutions concrètes d'« éthique *by design* » proposées traduisent des principes éthiques en normes et réglementations et produisent une *soft law* proposée aux entreprises et aux pouvoirs publics. Même si l'adhésion se fait sur une base volontaire, ce « droit mou » est appelé à prendre une dimension de plus en plus contraignante en se généralisant.

Dans son action normative, *IEEE* met en place un processus de consultations qui se veut très ouvert car il souhaite aboutir à des normes consensuelles. Cependant, la sociologie de l'organisation, ses méthodes, sa culture et les rapports de force⁽⁴⁹⁾

(44) « Konstantinos Karachalios (...) championed expansion of IEEE influence in key techno-political areas, including consideration of social and ethical implications of technology, according to the IEEE mission to advance technology for humanity » (www.ieee.org/about/management-council.html).

(45) « IEEE Standards Association Introduces Global Initiative for Ethical Considerations in the Design of Autonomous Systems », 5 avril 2016 (https://standards.ieee.org/news/2016/ieee_autonomous_systems.html).

(46) *Ethically Aligned Design—A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems (version 1 for public discussion)*, IEEE, 13 décembre 2016, p. 123 (<https://standards.ieee.org/>).

(47) HENNESSEY Meghan, *Clearpath Robotics Takes Stance Against 'Killer Robots'*, 13 août 2014 (www.clearpathrobotics.com/2014/08/clearpath-takes-stance-against-killer-robots/).

(48) La 1^{re} version du document (13 décembre 2016) (<https://standards.ieee.org/>) et la 2^{de} (<https://ethicsinaction.ieee.org/>).

(49) Financées par l'industrie, les universités américaines investissent dans la recherche sur l'éthique des développements de l'IA et cette force d'expertise très pluridisciplinaire domine naturellement le débat international ; voir « Carnegie Mellon University Announces K&L Gates Professorships », 03 avril 2018 (www.cmu.edu/).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

représentés font qu'une « main invisible » opère plutôt en faveur des intérêts de la recherche et de l'industrie américaine. D'autant que les acteurs industriels du domaine sont loin d'être passifs à l'image de Google, Amazon, Facebook, Microsoft et IBM qui annoncent, en septembre 2016, la création d'une organisation commune appelée « Partenariat pour l'IA » (*Partnership on Artificial Intelligence to Benefit People and Society* qu'Apple rejoint en janvier 2017) ⁽⁵⁰⁾.

Le 23 août 2017, deux ans après la très médiatisée « *Open Letter on AI* » co-signée par Stephen Hawking et les patrons des Gafami ⁽⁵¹⁾, une seconde lettre ouverte est présentée à Melbourne lors de l'*IJCAI-2017*. Le texte a été rédigé par les deux éminents professeurs britanniques Stuart Russell et Toby Walsh ⁽⁵²⁾. Pour ce dernier, il s'agit d'aller plus loin en menaçant de boycott les entreprises, mais aussi les universités qui participeraient à des projets militaires sur l'IA ⁽⁵³⁾. Le texte fait très directement référence aux travaux de la CCAC à Genève dont la première réunion du nouveau groupe d'experts gouvernementaux (GGE) sur les Sala vient d'être repoussée à la semaine du 13 novembre 2017 ⁽⁵⁴⁾.

Cette seconde étape inaugure un rôle plus visible d'Elon Musk qui a besoin de redorer une image quelque peu ternie par sa proximité avec Donald Trump depuis sa nomination au Conseil économique consultatif du Président ⁽⁵⁵⁾. Plus fondamentalement, cette initiative montre une volonté d'internationaliser, ou plus exactement de « dé-américaniser », la campagne en mettant en avant six signataires (sur 116) : Elon Musk (États-Unis), Mustafa Suleyman (Royaume-Uni), Esben Østergaard (Danemark), Jürgen Schmidhuber (Suisse), Yoshua Bengio (Canada, fondateur du *Montreal Institute For Learning Algorithms*) et le français Jérôme Monceaux (fondateur d'Aldebaran Robotics qui produit les robots *Nao* et *Pepper*).

Le choix des nationalités est très stratégique car chacun de ces pays a un rôle particulier à jouer dans la perspective d'une éventuelle négociation d'un « traité d'interdiction des armements autonomes ». L'Europe est clairement la cible principale identifiée, d'autant que la réflexion éthique institutionnelle, financée depuis de nombreuses années par l'Union européenne (UE), s'y est structurée autour d'une ligne plutôt pacifiste. Lors de l'*European Robotics Forum* qui se déroule en mars 2013 à Lyon, nous avons co-animé avec Noel Sharkey une table ronde où nous avons abordé cette impasse faite sur les sujets de la robotique militaire qui, aujourd'hui, ne permet pas à

(50) *The Partnership on AI to Benefit People and Society* (www.partnershiponai.org/).

(51) Google (Alphabet), Apple, Facebook, Amazon, Microsoft et IBM.

Demis Hassabis et Mustafa Suleyman pour Google-DeepMind, Steve Wozniak pour Apple ou Eric Horvitz pour Microsoft. (52) Toby Walsh est le principal rédacteur de la lettre (comme de celle de juillet 2015). Cette initiative coïncide avec le lancement de son livre *It's Alive! Artificial Intelligence from the Logic Piano to Killer Robots*, La Trobe University Press, juillet 2017, 369 pages.

(53) La première tentative de boycott est lancée en 2018 contre l'Institut supérieur coréen des sciences et technologies (KAIST) : VINCENT James, « Leading AI researchers threaten Korean university with boycott over its work on "killer robots" », *The Verge*, 04 avril 2018 (www.theverge.com/).

(54) Ce GGE créé lors de la conférence de révision de la CCAC en décembre 2016 et placé sous la présidence de l'Inde s'est réuni à deux reprises du 13 au 17 novembre 2017 et du 28 au 31 août 2018 (www.unog.ch/).

(55) Comme illustration de ce soutien voir MUSK Elon, « AI is vastly more risky than North Korea », *The Guardian*, 14 août 2017 (www.theguardian.com/technology/2017/aug/14/elon-musk-ai-vastly-more-risky-north-korea).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

l'UE de peser dans un débat international dominé par des voix britanniques et américaines. À l'époque, au sein de la Plateforme de technologie robotique européenne, le groupe d'experts sur l'éthique de la robotique rassemblait huit chercheurs : cinq Britanniques (dont trois de l'Université de Sheffield Amanda, Noel Sharkey et Tony Prescott) et trois Italiens. Cette même « géopolitique » se retrouve au sein du conseil scientifique du *Future of Life Institute (FLI)* qui, outre quelques chercheurs britanniques, compte aussi une Italienne : Francesca Rossi, professeur en informatique à l'Université de Padoue.

Les relais d'opinion que le *FLI* construit en Europe lui permettent de développer une stratégie de communication qui cible tout autant l'enceinte onusienne de la CCAC de Genève que les institutions européennes de Bruxelles. Ainsi, c'est d'un pays de l'Union européenne, la Suède, et avec l'appui de l'Association européenne pour l'IA (*EurAI*, ex-*ECAI*) que le *FLI* lance le 18 juillet lors de l'*IJCAI-2018* son « Engagement sur les Armes autonomes létales »⁽⁵⁶⁾. Ce texte appelle moins les gouvernements à signer un traité d'interdiction qu'« à créer un futur avec des normes internationales fortes, des règles et des lois contre les armes autonomes létales ». Cet engagement s'appuie sur la diffusion d'une vidéo de quelques minutes dans laquelle des essaims de microdrones dotés de capacités de reconnaissance faciale traquent et tuent de manière autonome des étudiants contestataires⁽⁵⁷⁾. Le film qui n'annonce que tardivement qu'il s'agit d'une fiction se termine par une intervention de Stuart Russell et affiche le lien autonomousweapons.org renvoyant à la campagne pilotée par le *FLI*. Il ne s'agit donc plus de soutenir la campagne *Stop Killer Robots* conduite par *Human Rights Watch*, mais d'affirmer le *leadership* du *Future of Life Institute* qui prend le contrôle du tempo médiatique et de la mobilisation internationale contre les armes autonomes.

Le 12 septembre 2018, la résolution commune sur les systèmes d'armes autonomes votée par le Parlement européen cite à deux reprises la lettre ouverte du *FLI* d'août 2017 « signée par 116 fondateurs d'entreprises de pointe dans les domaines de l'intelligence artificielle et de la robotique » qui semble avoir davantage impressionné les députés que l'action des ONG qui est simplement mentionnée parmi « les initiatives de la société civile », dans une liste qui commence par évoquer l'action du CICR⁽⁵⁸⁾.

Aujourd'hui : quelles avancées, quelles perspectives et quels écueils d'une surenchère transhumaniste ?

Cet article assume la subjectivité d'un observateur qui, au sein d'une administration, suit ce sujet depuis 2009. Nous éprouvons un profond respect pour les universitaires et chercheurs tels que Noel Sharkey, Jürgen Altmann ou Wendell Wallach qui alertent depuis plus de dix ans l'opinion publique internationale sur un sujet de légitime inquiétude. Notre regard est plus critique sur l'action d'une « industrie de

(56) *Lethal Autonomous Weapons Pledge* (<https://futureoflife.org/lethal-autonomous-weapons-pledge/>).

(57) LAM DX, *Killer robots, an increasingly real fiction* (www.youtube.com/watch?v=Sy-jyB2syx4).

(58) PARLEMENT EUROPÉEN, *Résolution sur les systèmes d'armes autonomes, 2018/2552(RSP)*, adoptée à Strasbourg le 12 septembre 2018 (www.europarl.europa.eu/).

la philanthropie » dont le souci de rentabilité de son investissement dans une campagne d'opinion pousse ces ONG à considérer comme secondaire l'objectif d'efficacité du mécanisme de contrôle des armements qu'elles proposent d'adopter. Mais cette critique serait encore plus forte à l'encontre d'industriels qui ont fait du *FLI* l'acteur principal du jeu du bonneteau évoqué en introduction.

En effet, même si le *FLI* affirme vouloir peser dans les débats à la CCAC, le véritable intérêt des industriels qui le financent se joue ailleurs et, en fait, ils ne portent qu'un intérêt limité au sujet des Sala. Pour nombre de ces sociétés de la Silicon Valley, il s'agit surtout d'une opération de diversion pour préserver l'image *cool* et altruiste de leurs activités. Après la controverse récente⁽⁵⁹⁾ sur l'implication de Google dans le projet *Maven* du Pentagone (alors que cette société affirme depuis des années n'avoir aucun lien avec le monde de la défense américain), on peut être assez circonspect sur la sincérité d'un engagement en faveur d'une « IA au service de l'humain – *AI for Humanity* »⁽⁶⁰⁾. Pour d'autres, comme Elon Musk ou l'Estonien Jaan Tallinn (Skype), leur motivation est sans doute plus sincère.

Ces entrepreneurs ont eu le mérite de replacer le débat vers la question centrale de l'objectif recherché, un pragmatisme qui conduit à penser davantage en termes d'encadrement que d'interdiction. Cette réflexion portant sur l'élaboration d'un modèle de gouvernance des progrès scientifiques était déjà présente au moment de la création de l'*ICRAC* et se retrouve encore dans la double recommandation – un peu schizophrène – qui apparaît fin 2012 en conclusion du rapport de *HRW* : *Losing Humanity*⁽⁶¹⁾. La plupart de ses lecteurs ont surtout retenu la première recommandation appelant « tous les États à interdire le développement, la production et l'utilisation d'armes totalement autonomes au travers de l'adoption d'un instrument de droit international juridiquement contraignant ». On n'a guère prêté attention à la seconde recommandation qui s'adresse spécifiquement « aux roboticiens et autres (chercheurs) impliqués dans le développement des armes robotisées ». Or, dans cet ultime paragraphe, il n'est plus question d'une interdiction totale des armes autonomes, mais du moyen d'encadrer et de contraindre leur production. Est ainsi proposé d'« établir un code de conduite professionnel orientant la recherche et le développement d'armes robotisées autonomes (...) de manière à garantir que les préoccupations juridiques et éthiques se rapportant à leur usage dans les conflits armés soient bien prises en compte à tous les stades du développement technologique ». Le texte se conclut sur l'exemple des « codes de conduite pour le développement de technologies militaires qui existent déjà dans les domaines de la biologie de synthèse ou des nanotechnologies ».

(59) Le projet *Maven* de « guerre algorithmique » utiliserait des données et des capacités d'IA de Google pour interpréter les images déterminant les frappes des drones américains. Plusieurs milliers d'employés de Google ont signé une pétition condamnant le « business de la guerre » ; SHANE SCOTT et WAKABAYASHI Daisuke, « "The Business of War": Google Employees Protest Work for the Pentagon », *The New York Times*, 04 avril 2018 (www.nytimes.com/).

(60) Titre de la journée du 29 mars 2018 au Collège de France présentant le rapport du député Cédric Villani et clôturée par l'annonce du président de la République d'une stratégie nationale sur l'intelligence artificielle. Au cours de cette journée se sont exprimés Yann LeCun, le responsable de Facebook AI Research, Demis Hassabis, le président de Google DeepMind ou encore le professeur de Berkeley Stuart Russell, trois des tout premiers signataires de la lettre ouverte de 2015 contre les armes autonomes.

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

Cette dualité des objectifs trouve davantage de cohérence dans les recommandations formulées en avril 2013 par le Rapporteur spécial sur les exécutions extrajudiciaires auprès du Conseil des droits de l'Homme de l'ONU qui invite « les États à appliquer un moratoire sur les essais, la production, l'assemblage, le transfert, l'acquisition, le déploiement et l'utilisation de robots létaux autonomes »⁽⁶²⁾, mais uniquement en attendant de se mettre d'accord sur les règles à appliquer pour « i/ l'acquisition, la détention et le stockage ; ii/ la recherche (fondamentale ou appliquée) et le développement ; iii/ les essais ; iv/ le déploiement ; v/ le transfert et la prolifération et vi/ l'usage »⁽⁶³⁾.

Tant l'ultime paragraphe du document de *HRW* que cette dernière citation du rapport de Christof Heyns se rapporte à une seule et même source : un article collectif de 36 pages paru en 2011 dans la *Columbia Science and Technology Law Review*⁽⁶⁴⁾. La dizaine d'auteurs qui y ont contribué a en commun d'avoir des liens directs ou indirects fréquents avec le monde de la défense américain et de réfléchir très concrètement à la manière dont les forces armées d'un pays démocratique peuvent utiliser les progrès technologiques sans renier ses valeurs. Mais de quelles valeurs morales parle-t-on et quel est le cadre de leur expression publique ?

Dans des sociétés qui font de moins en moins confiance à l'État ou aux organisations interétatiques pour exprimer l'intérêt général, certains proposent une alternative libérale, voire libertarienne. Ainsi, le *Partnership on AI (PAI)* lancé par les Gafami fin 2016 partage désormais sa gouvernance avec des représentants de la société civile et d'ONG investies dans les libertés publiques. Le *PAI* préfigure-t-il une forme souhaitable de « Nations unies de l'intelligence artificielle »⁽⁶⁵⁾ ? On peut s'inquiéter des intérêts mercantiles, mais aussi idéologiques qui demeurent derrière un engagement se présentant comme simplement « au service du bien » (*Don't be evil*, « Ne soyez pas malveillant », pour reprendre la devise de Google). On peut avoir les mêmes préventions face à une association professionnelle mondiale comme *IEEE* qui, malgré la diversité de ses 400 000 membres, peut être marquée par un parti-pris idéologique ou une vision du monde singulière.

Lors du « Sommet IA & Éthique » qu'*IEEE* a organisé à Bruxelles le 15 novembre 2016, le modérateur de l'événement était John C. Havens. Celui qui a été choisi par *IEEE* pour diriger sa *Global Initiative for Ethical Considerations in the Design of Autonomous Systems* présente un profil assez atypique. C'est un acteur professionnel qui a joué à Broadway ou pour des séries télévisées, rendu populaire par ses interventions dans les conférences *TED (Technology, Entertainment & Design)*

(61) *Losing Humanity*, *op. cit.*, p. 47-48.

(62) HEYNS Christof, *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*, ONU, 2013, 22 pages (www.ohchr.org/), cf. p. 21 (point n° 113).

(63) *Ibid.*, cf. p. 19 (point n° 103).

(64) MARCHANT Gary E., ALLENBY Braden, ARKIN Ronald C., BARRETT Edward T., BORENSTEIN Jason, GAUDET Lyn M., KITTRIE Orde F., LIN Patrick, LUCAS George R., O'MEARA Richard et SILBERMAN Jared, « International Governance of Autonomous Military Robots », *Columbia Science and Technology Law Review*, 30 décembre 2010 (<https://ssrn.com/abstract=1778424>).

(65) L'expression est de Mark NITZBERG, cité dans TUAL Morgane et PIQUARD Alexandre, « Éthique et intelligence artificielle : récit d'une prise de conscience mondiale », *Le Monde*, 04 octobre 2018.

diffusées sur YouTube. John C. Havens est le promoteur du « Projet H(app)apthon » qui plaide pour un monitoring technologique du sentiment de bonheur des individus afin d'orienter les politiques publiques ⁽⁶⁶⁾. Même s'il est entouré de personnes sérieuses ⁽⁶⁷⁾, on peut nourrir quelque inquiétude de le voir aujourd'hui à la tête d'un projet de réflexion éthique au sein de l'IEEE.

Avec le centre de recherche *OpenAI* de San Francisco, le *Center for Human-Compatible AI (CHAI)* de l'Université de Berkeley ou le *Future of Humanity Institute* dirigé par le philosophe Nick Boström à l'Université Oxford, c'est tout un univers de recherche qui gravite autour du *Future of Life Institute*. Cette réflexion est fortement teintée d'une idéologie transhumaniste qui, dans l'espoir de créer une « IA générale », peut inquiéter par la distance prise avec les valeurs politiques et philosophiques qui fondent notre modèle de démocratie libérale et la valeur de dignité humaine qu'elle promeut. Le mouvement intellectuel qui s'est désormais greffé au milieu de la recherche en sciences et technologies du numérique rend le débat public extrêmement caricatural d'autant que le financement de cette recherche repose en grande partie sur une logique de médiatisation.

Il est difficile aujourd'hui de faire la différence entre la vidéo *Slaughterbots* produite par Stuart Russel pour promouvoir la campagne « *Ban Autonomous Weapons* » ⁽⁶⁸⁾ et la bande-annonce d'un épisode de la série télévisée britannique *Black Mirror* ⁽⁶⁹⁾ ! On peut tout à fait imaginer retrouver les mêmes conseillers techniques travaillant à la fois pour Netflix ou le *FLI*. Cette production médiatique se rejoint en tout cas dans un même sentiment de défiance vis-à-vis d'une action publique perçue comme incapable de défendre l'intérêt général ou de porter un idéal universel au service de l'humanité.

Pourtant, si l'on souhaite insuffler de l'éthique dans les développements des technologies émergentes dans le domaine de la défense, il est nécessaire de revaloriser l'action publique ceci afin de dégager la réflexion d'un cadre qui ne se voudrait qu'émotionnel (laisser l'« effet “beurk” » guider notre boussole morale comme nous y invite Jody Williams et *HRW*). Cette réflexion éthique et sociétale ne doit pas être monopolisée par certains intérêts nationaux ou commerciaux, mais chaque pays doit pouvoir reprendre une certaine maîtrise de ses choix technologiques et du modèle de gouvernance de leurs usages. Il ne s'agit pas de se refermer sur soi, mais au contraire d'être plus serein pour s'ouvrir à un dialogue multilatéral indispensable pour être efficace. Comme l'a rappelé le ministre de la Défense dès l'été 2013, l'objectif est bien d'assurer « l'universalité d'un éventuel nouvel instrument de droit international » car chaque exemple de sortie du cadre de la CCAC (les processus d'Oslo ou d'Ottawa) a conduit à exclure « les plus grandes puissances militaires (États-Unis, Russie, Chine, Inde, Pakistan, Israël...) ». Il en est de même aujourd'hui pour les Sala dont tout futur

(66) (<http://happathon.com/>) Il est aussi l'auteur de l'ouvrage *Heartificial Intelligence: Embracing Our Humanity to Maximize Machines*, TarcherPerigee, 2016, 304 pages.

(67) Dont les professeurs Raja CHATILA et Catherine TESSIER qui participent aux travaux de l'IEEE et ont coécrit un article particulièrement utile « A Guide to Lethal Autonomous Weapons Systems », 04 septembre 2018, publié sur le site du CNRS (<https://news.cnrs.fr/opinions/a-guide-to-lethal-autonomous-weapons-systems>).

(68) *Slaughterbots*, 7 min 47 (<https://autonomousweapons.org/slaughterbots/>).

(69) On peut penser à l'épisode n° 6 de la 3^e saison, intitulé « Hated in the Nation ».

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

encadrement ne peut avoir de valeur que s'il est adopté par les Nations capables de développer ce type de technologies. Le texte que la France et l'Allemagne ont conjointement proposé au groupe d'experts gouvernementaux sur les Sala a précisément pour objet de rechercher un tel consensus auprès des États parties à la CCAC ⁽⁷⁰⁾. Seul un accord très large pourra garantir la robustesse d'un processus de contrôle d'usages contestables de l'IA et limiter le risque de retrouver ces technologies entre les mains d'acteurs non étatiques affranchis de tout scrupule.

*
**

Le processus diplomatique conduit à Genève est long et certainement frustrant à bien des égards. Il peut prendre encore quelques années avant de porter ses fruits, mais où se trouve la réelle urgence aujourd'hui ? L'exercice de la responsabilité politique implique de savoir hiérarchiser les enjeux or, pour le moment, les problématiques soulevées par la robotisation du champ de bataille relèvent moins de l'univers physique que du domaine informationnel. Si la guerre a été définie par Clausewitz comme « un acte de violence destiné à contraindre l'adversaire à réaliser notre volonté » ⁽⁷¹⁾, les technologies numériques ouvrent désormais la possibilité d'une soumission des esprits « sans violence ». Avant de tuer des êtres humains, la première victime de ces applications technologiques pourrait être nos démocraties elles-mêmes. L'urgence est bien là !

Pour la France et l'Europe, il est impératif de maintenir une réflexion éthique structurée et permanente, qui n'oppose pas l'action publique et la mobilisation de la société civile, l'IA militaire et l'IA civile. Cette réflexion éthique conduite aux États-Unis depuis de nombreuses années a su s'appuyer sur un débat public qui a mis en œuvre des moyens financiers et humains conséquents pour aboutir à des recommandations techniques répondant aux exigences morales d'une démocratie libérale ⁽⁷²⁾. Dès lors, en France ou en Europe, avons-nous encore besoin de nous poser des questions, puisque les réponses nous sont déjà offertes « clef en main » ?

En parlant d'« autonomie » de la machine, on oublie celle de l'homme et de la société dans laquelle il inscrit son destin. L'autonomie d'une société porte un nom, c'est sa souveraineté, un concept qui est profondément bouleversé par les progrès des technologies du numérique ⁽⁷³⁾. La souveraineté dans un débat éthique c'est la liberté de se poser des questions et d'y répondre mais aussi, plus en amont, la liberté de

(70) Déclaration commune franco-allemande « For consideration by the Group of Governmental Experts (GGE) on Lethal Autonomous Weapons Systems (LAWS) – CCW/GGE.1/2017/WP.4 » présentée le 07 novembre 2017 (www.undocs.org/ccw/gge.1/2017/WP.4).

(71) « Der Krieg ist also ein Akt der Gewalt, um den Gegner zur Erfüllung unseres Willens zu zwingen », CLAUSEWITZ (VON) Carl, *Vom Kriege*, 1832, chapitre 3 (<http://gutenberg.spiegel.de/buch/vom-kriege-4072/3>).

(72) KISSINGER Henry, « How the Enlightenment Ends », *The Atlantic*, juin 2018 (www.theatlantic.com/) ou la récente profession de foi de BOYD Aaron, « Pentagon Doesn't Want Real Artificial Intelligence in War, Former Official Says », *Defense One*, 31 octobre 2018 (www.defenseone.com/).

(73) GANASCIA Jean-Gabriel, GERMAIN Éric et KIRCHNER Claude, *La souveraineté à l'ère du numérique : Rester maîtres de nos choix et de nos valeurs*, Cerna, octobre 2018 (<http://cerna-ethics-allistene.org/>).

Trouvez le Sala !

La campagne internationale contre les « robots tueurs » : une décennie et déjà trois histoires

formuler ces questions. La bonne IA n'est pas nécessairement celle qui est validée au sein de l'*IEEE P7000™ Standards Working Group* ⁽⁷⁴⁾ et la France, comme l'Europe, doit aujourd'hui prendre la responsabilité de développer son propre modèle de gouvernance éthique de l'intelligence artificielle ⁽⁷⁵⁾. ♦

(74) Le groupe de travail « Model Process for Addressing Ethical Concerns During System Design—IEEE P7000™ » présidé par John C. Havens (<https://standards.ieee.org/>).

(75) Nous reprenons à notre compte la conclusion du rapport de synthèse du débat public animé par la Commission nationale de l'informatique et des libertés (Cnil) sur les enjeux éthiques des algorithmes et de l'IA : *Comment permettre à l'Homme de garder la main ? Rapport sur les enjeux éthiques des algorithmes et de l'intelligence artificielle*, 15 décembre 2017 (www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf).

