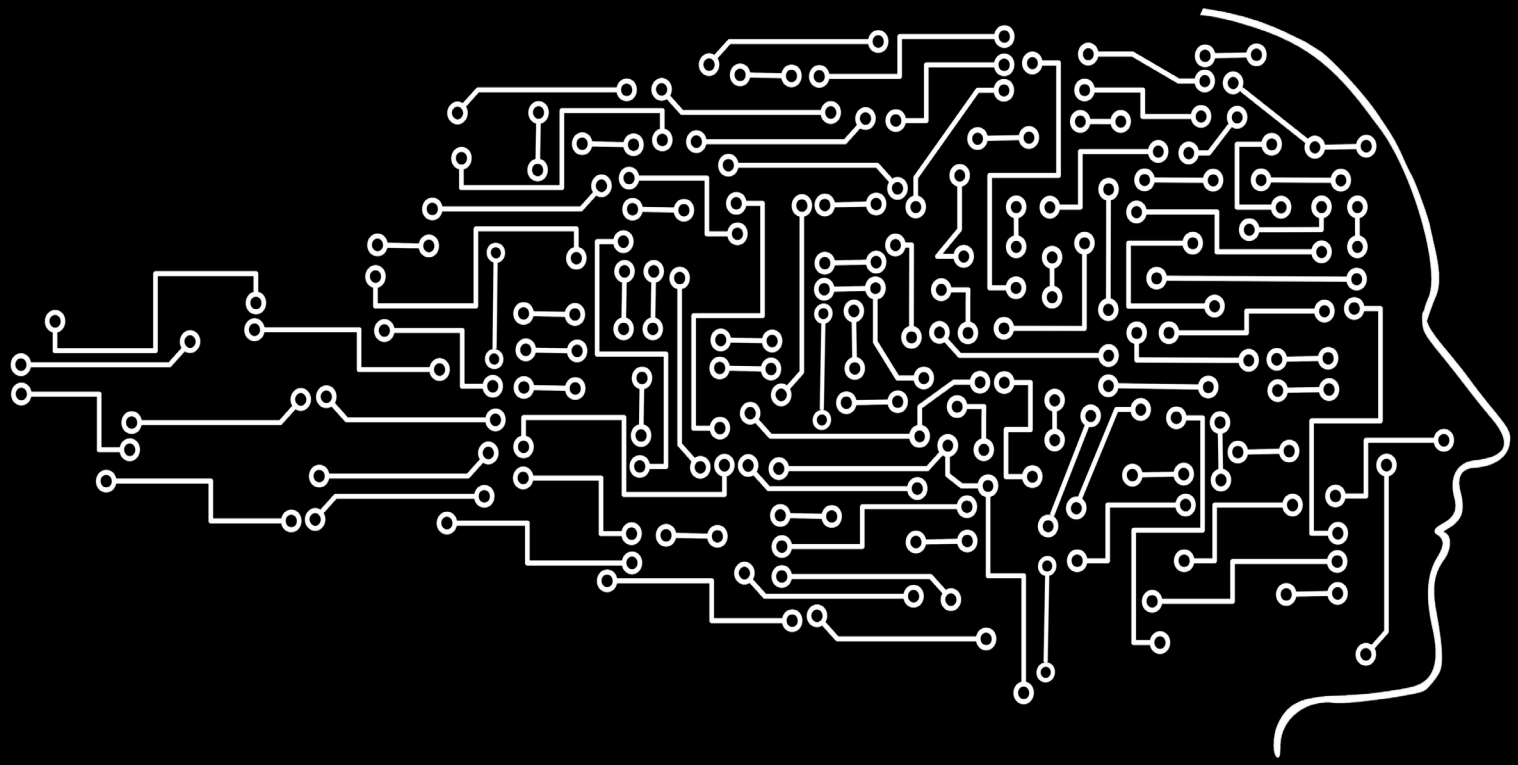
SÉCURITÉ & STRATÉGIE 146



Les organisations de défense face aux défis de l'intelligence artificielle

Alain DE NEVE

INSTITUT ROYAL SUPERIEUR 150 ans think tank de DEFENSE

Mars 2021

CENTRE d'ÉTUDES de SÉCURITÉ et DÉFENSE
SÉCURITÉ & DÉFENSE

Les organisations de défense face aux défis de l'intelligence artificielle

Résumé

La présente étude a pour but d'examiner l'impact des technologies de l'intelligence artificielle (IA) dans les différentes dimensions du champ stratégique et de l'action militaire. Comme nous nous emploierons à le démontrer, les technologies constitutives de l'IA demeurent encore mal comprises. Ceci est la conséquence de l'imaginaire qui a entouré cette technologie et fait d'elle le cœur d'une véritable « mystification ». La portée des transformations qu'elles englobent sur le plan de l'agir militaire sont mal identifiées et sous-estimées. Or, les technologies de l'intelligence artificielle ne bouleverseront pas seulement les moyens d'action des organisations de défense qui parviendront à les acquérir ; elles affecteront comme jamais auparavant la perception même de l'environnement stratégique par les décideurs, de même que leurs procédures de décision.

La nouvelle course aux technologies de l'IA d'ores et déjà engagée entre plusieurs puissances scientifiques, technologiques et militaires ne fera que s'intensifier et ce, dans un monde caractérisé par une dissolution progressive des régimes multilatéraux de sécurité. Des horizons conceptuels inexplorés se présenteront à moyen-terme aux forces armées appuyées par l'IA : analyse en temps réel de l'environnent stratégique basée sur le traitement de bases de données toujours plus étendues, surveillance tous azimuts des signaux faibles de l'environnement stratégique, puissances de calcul au service d'un tempo opérationnel susceptible d'échapper à tout supervision humaine, processus de décision assistés par l'IA hors de portée de l'intelligence humaine, armements et plates-formes équipés de dispositifs de maintenance prédictive afin de réduire les latences logistiques, etc.

À l'aube des transformations qui s'annoncent et se précisent, les implications pour la Belgique, l'OTAN et l'Union européenne s'avèrent considérables. Les technologies de l'IA appellent non seulement à des investissements nouveaux dans des secteurs technologiques disruptifs mais aussi à l'établissement de normes nouvelles, de cadres de régulation et de mesures de confiance et de sécurité prenant en compte toute la spécificité de l'IA. La connaissance et l'encadrement normatif ne peuvent être dissociés. Elles représentent des réalités complémentaires. Par ailleurs, la complexité et l'étendue des champs d'action concernés par les technologies de l'IA appellent à des décisions et des actions concertées sur le plan multilatéral, que ce soit à l'échelle de l'OTAN ou de l'Union européenne. L'investissement de la Belgique et de ses partenaires concerneront donc des secteurs aussi divers que l'enseignement et la formation scientifique et technologique, la transition industrielle et le soutien aux hautes technologies, l'accès aux connaissances et processus essentiellement maîtrisés par des groupes technologiques privés qui se situent à l'origine des principales percées de l'IA depuis 20 ans. L'évitement d'une course aux technologies de l'IA aux effets déstabilisateurs exigera la formulation de stratégies concertées sur le plan multilatéral mais aussi l'assurance que l'ensemble des pays qui prendront part à une telle entreprise de concertation dispose et maîtrise les connaissances nécessaires à la régulation des technologies constitutives de l'IA.

INSTITUT ROYAL SUPERIEUR de DEFENSE
150 ans think tank

SÉCURITÉ & STRATÉGIE 146

Mars 2021

CENTRE d'ÉTUDES de SÉCURITÉ et DÉFENSE
SÉCURITÉ & DÉFENSE

À propos de l'auteur

Politologue de formation, titulaire d'un diplôme d'études approfondies en sciences politiques (orientation : relations internationales) de l'Université catholique de Louvain, Alain De Neve est chercheur spécialisé dans les questions relatives aux développements technologiques en matière de défense au sein du Centre d'études de sécurité et défense (CESD). Plus spécifiquement, Alain s'attache à l'étude de l'innovation technologique en matière de défense, à l'industrie de l'armement ainsi qu'au domaine aérospatial.

Dans le cadre de ses activités au sein du CESD, Alain De Neve se consacre pleinement aux problématiques liées à l'innovation technologique, aux technologies émergentes et à l'évolution du marché européen des biens et équipements de défense. Spécialiste de ces multiples sujets, il intervient comme consultant dans de nombreux forums et médias belges et étrangers. Alain De Neve collabore également à divers projets de publication académiques de l'Université libre de Bruxelles et de l'Université catholique de Louvain. Enfin, il est régulièrement invité par des universités et associations à dispenser des formations et donner des conférences sur les problématiques globales de sécurité en lien avec la technologie.

« *La stratégie est l'art de faire la guerre intelligemment* »[1]

Général Claude Le Borgne

« *Ce qui peut être fait techniquement, le sera nécessairement* »[2]

Dennis Gabor

[1] Claude LE BORGNE, *La Guerre est morte, mais on ne le sait pas encore*, Grasset, 1987.

[2] Dennis GABOR (1900 – 1979) était un ingénieur et physicien hongrois qui reçut le prix Nobel de physique en 1971 pour l'invention de l'holographie. Dennis Gabor développa au cours de sa carrière une vision critique de la technologie de son époque. Affirmant que la technologie possède ses propres lois d'expansion et d'évolution, il peut être, d'une certaine façon, rattaché au courant *déterministe*. Le développement technologique était selon lui le résultat de deux phénomènes : la constante recherche d'innovations de la part de l'industrie technique, destinées à la maintenir dans son existence, et les conséquences de la logique même de la société technicienne apparue au lendemain de la Seconde Guerre mondiale.

Table des matières

Introduction

La révolution copernicienne fut vécue par l'homme de l'époque comme un réel bouleversement cosmologique. La Terre, que l'orgueil de ce dernier avait placée au centre de son univers, n'était finalement qu'une planète comme les autres occupant une place tout à fait quelconque dans un système solaire d'une extraordinaire diversité. De l'immobilité découlant de cette prétendue centralité, la Terre et l'homme basculaient dans le mouvement et avec elle dans l'incertitude du changement perpétuel et incontrôlé. Plus tard, la révolution darwinienne allait dévaloriser l'homme une fois de plus en lui signifiant qu'il n'était sur cette Terre qu'une espèce parmi tant d'autres dans la richesse du vivant. La domination qu'il pensait exercer sur le règne animal et végétal n'était finalement qu'illusion, tant la complexité du vivant échappait à son observation directe et à sa compréhension. Darwin démontrait par ailleurs qu'*homo sapiens sapiens* n'était que le résultat de millions d'années de combinaisons que le hasard de la nature avait laissé émerger. Ce que l'on pensait alors être le coup de grâce vint lorsque Freud déconstruisit la maîtrise que l'homme pensait avoir de lui-même en lui expliquant qu'il était en bonne part gouverné par un principe sur lequel il n'avait guère de prise : l'inconscient. La seconde moitié du XX^{e} siècle, débutant avec Hiroshima et Nagasaki, se poursuivant avec la menace d'une destruction mutuelle assurée (MAD) par échanges balistiques thermonucléaires et s'ancrant avec le mythe d'une croissance économique sans freins ni garde-fous, redonna temporairement l'illusion de domination à l'homme. Mais le XXI^{e} siècle pourrait faire surgir une nouvelle remise en cause ontologique du statut de l'homme. Peut-être, selon certains, s'agira-t-il là de son ultime revers : non seulement l'homme n'est maître de rien, mais il s'apprête en outre à faire émerger les conditions de son futur asservissement par un artefact issu de son propre génie : l'intelligence artificielle.

Des machines intelligentes remplaceront-elles un jour le soldat dans la conduite de la guerre ? La décision d'engagement des forces, qu'il s'agisse de soldats ou de systèmes d'armes, dépendra-t-elle demain de dispositifs issus de l'intelligence artificielle ? La guerre, « l'art de la dialectique des volontés », selon les termes d'André Beaufre, soumise aux passions humaines et multiples biais cognitifs qui affectent les processus de décision, est-elle en passe de s'effacer à la faveur de processus analytiques froids et complexes… tellement complexes que l'homme ne serait plus en mesure de décrypter l'enchaînement des données et de leur traitement qui auront conduit à l'engagement d'hostilités ? Ce sont là quelques exemples, certes parfois fantasmés, des multiples interrogations qui portent sur la place qu'occupent et qu'occuperont plus encore demain les systèmes d'IA[3] dans la conduite des opérations militaires. Avant de mesurer la place des systèmes d'IA au sein des dispositifs de défense des Etats qui ont fait le choix de développer – ou d'acquérir – et d'intégrer les technologies qui leur sont associées, encore faut-il comprendre ce que l'on désigne par l'IA. Or, le sens commun a fini par travestir cette notion : ce que l'on désigne, aujourd'hui, sous le label d'IA fait rarement référence à son emploi originel.

[3] Lorsque nous parlerons des technologies d'intelligence artificielle, nous emploierons préférentiellement l'expression « systèmes d'IA » et non le terme « IA ». En effet l'« IA », comme nous le verrons dans la suite de cet exposé, est une discipline scientifique, et non ce que le langage commun en fait, à savoir une expression désignant des systèmes développés dans le cadre de la discipline qu'est l'intelligence artificielle.

1. Pas un jour sans IA…

Une simple recherche à propos de l'expression « intelligence artificielle » sur Internet suffit à mesurer l'ampleur avec laquelle l'IA a envahi, en moins de dix ans, l'ensemble des secteurs de la société, en particulier le monde médiatique. L'IA constituerait tout à la fois la solution clé pour notre avenir et le pire des scénarios à envisager pour le devenir de l'Humanité. L'IA serait, au vu des intitulés des articles de presse (qu'elle soit généraliste ou de vulgarisation scientifique), la panacée à une grande diversité de problèmes et de défis : détection des cas de cancer les plus furtifs, connaissance approfondie et plus affinée du dérèglement climatique, appui aux forces de l'ordre, assistance juridique, aide à l'enseignement, biométrie, moteurs de recherche sur l'internet, connaissance de l'univers, exploration du vivant, etc. Il n'est pas un secteur qui ne soit impacté par les vertus de l'IA. À l'évidence, la question de la place de l'IA dans notre monde se doit d'être posée dans des termes plus mesurés et certainement plus dépassionnés.

En moins d'une décennie, l'intelligence artificielle s'est donc imposée comme une rupture technologique incontournable pour une immense variété de systèmes, et ce dans un large spectre de secteurs d'activités. L'intelligence artificielle, pourtant, est bien plus ancienne. Elle est surtout, à son origine, une discipline scientifique qui fut portée par des théoriciens et ingénieurs issus de la dynamique scientifique de la Seconde Guerre mondiale aux États-Unis. Elle est surtout le résultat d'investissements militaires considérables.

Sur le plan géopolitique, elle est devenue un enjeu prioritaire de défense pour nombre de puissances à travers la planète. Les États-Unis et la Chine sont, sans nul doute, les deux principaux acteurs qui ont considérablement investi dans un large éventail de programmes ayant pour objectif d'accroître la qualité des algorithmes des systèmes d'IA. Après eux, les États européens et la Russie semblent à la marge des percées technologiques en la matière. Il est essentiel de comprendre l'ampleur de la rupture sociétale sur le point de survenir avec l'irruption de l'IA dans l'ensemble des activités humaines. D'une certaine façon, l'IA est déjà présente dans une extraordinaire variété de domaines de production, de consommation et de modalités de gestion. Il est, aujourd'hui, plus difficile que jamais d'anticiper les transformations sociales et politiques qui découleront de l'immixtion croissante de l'IA au sein des communautés humaines. La célérité avec laquelle de telles transformations apparaissent conduit même les experts les plus éminents et réputés de la planète à alerter l'opinion sur les dangers que pourrait faire courir à l'Humanité entière la combinaison future entre l'IA, la robotique, les nanotechnologies et les biotechnologies. Risque de fracture, tout d'abord, entre deux formes d'intelligence – humaine et informatique –, la première étant désormais, selon certains, physiologiquement incapable d'appréhender les performances des machines. Risque, ensuite, d'une course nouvelle aux armements (d'ores et déjà engagée du reste) entre les puissances technologiques parvenues à maîtriser la conception de systèmes d'IA et les autres pays, relégués à l'arrière-plan des relations internationales et subissant les décisions de politiques étrangères et militaires développées sur la base des conjectures projetées par des systèmes d'IA. Au sein même des sociétés les plus avancées existe aussi le spectre d'une opposition entre ceux qui détiendront l'expertise dans la programmation et la compréhension des mécanismes sous-jacents de l'IA et ceux qui se verront exclus de la connaissance de tels mécanismes pour ne devenir que les « sujets » des programmations conçues.

En tant que discipline, l'IA – nous l'avons dit – n'en est pourtant pas à ses débuts et les résultats auxquels elle aboutit aujourd'hui sont le fruit de nombreuses années de recherche dont les prémices remontent aux années 1950. Longtemps marqué par une succession de cycles évolutifs et « involutifs » propres à tout développement technologique, le secteur de l'intelligence artificielle a récemment prospéré du fait de l'implication d'une multitude d'acteurs issus du secteur privé et, plus précisément, de ce que l'on pourrait qualifier d'« entrepreneuriat numérique ». Des sociétés telles que Google,

Amazon, Apple, Facebook mais aussi, en Asie, Huawei, Xiaomi, Samsung ont porté les performances de l'IA a des niveaux inégalés jusqu'alors. La concentration de l'expertise en matière d'IA entre les mains de ces acteurs d'un type nouveau n'est pas sans poser de questions sur les modalités de partage de cette technologie ou encore sur la capacité des pouvoirs publics à disposer d'un accès aux connaissances qui se situent à la base de l'IA. Nous aurons l'occasion d'aborder plus longuement cette question dans ces pages. Pourtant, la vitrine civilo-commerciale de l'IA contemporaine ne doit pas nous distraire d'une réalité fondamentale : l'intelligence artificielle, qu'on l'appréhende comme discipline ou comme produit technoscientifique, est avant tout le produit de la recherche militaire. L'IA est née pour des besoins militaires, sert aujourd'hui des besoins militaires et appuiera, demain encore, des besoins militaires. Passer à côté de cette réalité reviendrait à méconnaître les enjeux de la compétition internationale aujourd'hui lancée dans le domaine de l'IA. Nous observerons, par ailleurs, que cette compétition présente des particularités qui la distinguent des précédentes formes de course aux armements, qu'il s'agisse des armes conventionnelles ou des armes non-conventionnelles (en particulier le champ de la dissuasion nucléaire, avec laquelle l'IA est souvent comparée par nombre d'experts). Que l'on considère l'IA comme le cœur d'une nouvelle course aux armements ou qu'on l'envisage comme le « simple » accélérateur d'une course militaire, elle est de toute évidence associée aujourd'hui à l'expression moderne de cette recherche perpétuelle de systèmes d'armes performants capables de renverser les équilibres militaires.

Il n'existe pas aujourd'hui un domaine d'activité qui ne soit concerné par les progrès intervenus dans le champ de l'intelligence artificielle ou, plus exactement, *des* intelligences artificielles. Il pourrait même être affirmé sans peine qu'une majorité des dispositifs techniques sur lesquels nos activités quotidiennes s'appuient – que ces activités soient d'ordre privé ou professionnel – dépendent en grande partie d'intelligences artificielles spécifiques. Nous ne pourrions aujourd'hui concevoir nos déplacements, nos modes d'acquisition de l'information et des connaissances sans le recours à l'IA. À l'instar de nombreux domaines, le champ militaire n'est pas épargné par l'irruption de l'intelligence artificielle. Mais encore faut-il savoir de quoi l'on parle lorsque l'on évoque l'intégration de l'IA dans le domaine militaire et des armements. Ce n'est qu'au terme d'une compréhension correcte de ce que constitue le rôle que peut jouer l'IA dans le secteur militaire que nous pouvons déduire les corolaires techniques, politiques et éthiques d'une telle association. Pour l'exprimer autrement, la détermination des implications éthiques de l'IA dans le champ militaire suppose au préalable de nous interroger sur le statut de cette technologie dans la progression matérielle de nos systèmes de force.

2. Un concept galvaudé

Analyser la place qu'occupe aujourd'hui l'IA dans le champ militaire est une démarche exigeante à plus d'un titre. Tout d'abord, un examen sérieux de la problématique impose le dépassement de nombreux malentendus et fantasmes en ce qui concerne ce vaste ensemble de technologies. En tant que discipline de recherche, l'IA éprouve d'importantes difficultés à se détacher des mythes que l'imaginaire commun, notamment alimenté par les visions véhiculées par la science-fiction (principalement hollywoodienne), a construit au fil de quelques décennies. Mieux comprendre en quoi l'IA peut – ou pourra – influer sur la conduite de l'action stratégique exige de dissocier celle-ci des innombrables utopies et dystopies qui lui sont trop souvent associées. C'est sans doute là que se situe l'un des premiers obstacles à une juste appréhension de cet objet technique. Cela ne signifie pas pour autant que l'imaginaire de la science-fiction et des récits d'anticipation soit totalement dénué d'enseignements ou de réflexions à propos des enjeux de l'IA. On peut même considérer que certains récits posent de réelles questions sur le statut des rapports futurs entre l'homme et la machine ou sur la place qu'occuperont demain les systèmes synthétiques de décision dans la société. Que l'on évoque les trois lois de la robotique d'Asimov ou la problématique de l'omniprésence de la technique et des rapports humains/synthétiques comme le fit Philip K. Dick, de nombreuses narrations mettant en jeu

des formes d'intelligence artificielle dans nos sociétés intègrent des réflexions qu'il serait regrettable de balayer d'un simple revers de la main.

Comprendre les enjeux pratiques de l'IA dans l'ensemble des composantes de nos sociétés exige encore de nous affranchir des visions portées par un courant « évangéliste technologique » particulièrement présent aux États-Unis. Structuré autour d'un certain nombre de prophéties auto-réalisatrices, ce discours de type scientiste est porté par plusieurs éminents spécialistes et experts impliqués dans la recherche en matière d'intelligence artificielle et dans le domaine des algorithmes avancés. Parmi eux, on mentionnera tout particulièrement Ray Kurzweil, dont la plume prolifique a permis de porter le sujet des rapports entre l'IA et l'humain au cœur du débat politique et scientifique outre-Atlantique. Leader du courant transhumaniste, Ray Kurzweil affirme ni plus ni moins que l'avènement d'une fusion entre l'homme et la machine est en passe de devenir une réalité dans les décennies à venir. L'approche transhumaniste portée par ces techno-évangélistes auxquels appartient Kurzweil, sous ses dehors révolutionnaires, n'apporte en soi pas de véritable nouveauté conceptuelle en ce qu'elle affirme que la technique serait désormais en mesure de transformer l'homme, de l'améliorer pour lui permettre de dépasser la condition même de son humanité. Ainsi que le rappelle Jean-Michel Besnier, « *l'idée que les techniques ont la faculté de transformer l'homme est loin d'être neuve. Elle est au principe de la paléoanthropologie et de l'explication du phénomène de l'hominisation*. »[4] La prétention de Ray Kurzweil et de ses disciples réside toutefois dans un argument supplémentaire : celui selon lequel le pouvoir actuel de la technique aurait généré une forme de « coupure » dans le processus par lequel l'homme a justement produit ces techniques[5]. En d'autres termes, la nature même de la technique qui se déploie aujourd'hui sous nos yeux ne permettrait plus de préserver les conditions mêmes de l'humanité qui l'a originellement produite[6]. Cette réflexion n'est pas sans conséquence sur le statut futur du combattant au sein des organisations militaires. La question est souvent posée de savoir si les technologies issues des ordinateurs et des réseaux (produits de la révolution informatique des années 1970) sont appelées à transformer les modes de conduite de la guerre ou la nature de la guerre elle-même. Selon que nous contestons ou que nous nous rallions à la thèse transhumaniste, les deux seules réponses possibles à cette question s'avèrent diamétralement opposées. Affirmer que l'essence de la technologie n'est pas seulement génératrice de modalités nouvelles de combat mais productrice d'une transformation en profondeur de la nature de la guerre (non plus conçue comme une lutte dialectique entre des volontés humaines mais entre des intelligences synthétiques et humaines ou entre intelligences synthétiques seules) et, a fortiori, de la nature du combattant revient à nous projeter vers un horizon d'enjeux fondamentalement inédits. Nous examinerons plus en détail la portée philosophique de ces questions et leurs conséquences pour l'institution militaire dans la suite de ce travail.

Une étude sur l'apport de l'IA dans le champ militaire suppose aussi de toucher aux enjeux socio-organisationnels des organisations de défense. Comme toute grande organisation, le milieu des forces armées comporte des rapports humains articulés par des systèmes hiérarchiques, certes, mais aussi des relations de pouvoir et d'influence. Que l'on considère l'IA comme un adjuvant technique à l'accomplissement de certaines fonctions militaires ou comme un outil de remplacement de l'homme dans son autonomie décisionnelle, l'IA s'apprête à bousculer les modèles existants des rapports qui

[4] Jean-Michel Besnier, « Les nouvelles technologies vont-elles réinventer l'homme ? », *Études*, 2011, volume 6, tome 414, pp. 736-772.

[55] Jean-Michel Besnier, *Demain, les post-humains. Le futur a-t-il encore besoin de nous ?*, Paris, Fayard, 2020.

[6] Le courant transhumaniste envisage cette hypothèse dans une perspective résolument optimiste en affirmant que l'homme est enfin en mesure de dépasser ses limites physiques grâce à l'avènement de machines dont le niveau inégalé d'intelligence (porté par des capacités de calcul pratiquement sans limites) auquel il serait associé signifierait que le destin de l'Humanité ne serait plus scellé à sa condition terrienne. Cet espoir auquel travaillent les transhumanistes est incarné, depuis 2000, par l'Institut de la singularité (The Singularity Institute), créé par Ray Kurzweil avec des fonds apportés par Google et la NASA.

organisent la division du travail au sein des forces armées. Comme nous aurons l'occasion de le détailler dans la suite de ce travail, l'apport réel de l'IA en matière militaire s'illustrera moins sur le théâtre opérationnel que dans l'aide à la prise de décision, voire dans la prise de décision elle-même. Autrement dit, et au risque de verser dans une image quelque peu grossière, l'IA aura davantage les dehors d'un général que d'un fantassin. Des interrogations demeurent quant à l'acceptabilité d'un semblable scénario. Un soldat acceptera-t-il qu'une intelligence synthétique lui transmette des ordres ? Considérera-t-il ceux-ci comme légitimes ? Sera-t-il davantage enclin à une remise en question de la validité des consignes qui lui seront transmises par l'IA ? À l'autre bout de la chaîne de décision, le politique consentira-t-il à céder partiellement ou intégralement la conduite d'une action militaire à une machine ? Les responsables politiques et militaires consentiront-ils à confier une partie – sinon l'ensemble – du processus opérationnel à un système dont les clés de fonctionnement s'avèrent maîtrisées par un groupe restreint d'experts qui, dans le monde, comprennent le fonctionnement de la « boîte noire » d'un système d'IA ?

Une autre difficulté à laquelle est confronté quiconque examine le rôle possible de l'IA dans le champ militaire réside dans la confusion communément entretenue entre IA et armes autonomes. Il est généralement supposé que le rôle de prédilection de l'IA en matière de défense consistera à doter les armements d'une intelligence synthétique embarquée pouvant les rendre capables d'un pouvoir de décision. Comme nous le verrons dans la suite de ce travail, une telle hypothèse relève, à bien des égards, sinon du fantasme, à tout le moins d'une méconnaissance des possibilités et limites des technologies de l'IA. Il est pourtant courant de constater, y compris parmi les observateurs les plus avisés (beaucoup plus rarement parmi les praticiens militaires), ce type d'amalgame. Drones, armes autonomes, IA : il s'agit là de réalités qui entretiennent certes un rapport entre elles mais qui doivent être étudiées de manière distincte afin de définir exactement la portée de chacun de ces concepts.

Évaluer ce que pourrait être le rôle de l'IA dans le champ militaire suppose par ailleurs de procéder à l'examen d'un champ praxéologique qui, dans l'état actuel du développement technologique, n'existe pas en tant que tel.

À l'instar de toute nouvelle technologie, l'IA a ses partisans et ses détracteurs. Toutefois, il convient de dépasser leurs discours pour prendre la juste mesure de la réalité de cette technologie, ses atouts et ses inévitables limites. Une application d'IA est d'abord un modèle mathématique, entraîné avec des données numériques considérées comme représentatives de l'environnement dans lequel la machine agira ensuite. Ceci pose donc comme première limite la mise à jour régulière des données en fonction de l'évolution du cadre d'emploi. Mais, au-delà des applications informatiques, c'est l'ensemble des organisations qu'il faut adapter en vue d'une exploitation optimale de l'IA. Ceci suppose une réflexion et une réforme approfondie des infrastructures et processus de collecte, de stockage et de partage des données, des ressources humaines et des compétences qui lui seront consacrées.

La présente étude ne constitue nullement un panorama exhaustif des applications militaro-techniques de l'intelligence artificielle. Comment le pourrait-elle face au foisonnement quotidien d'éditoriaux, d'articles, d'études, de rapports, d'ouvrages consacrés aux divers domaines dans lesquels l'IA, y explique-t-on, est appelée à étendre son règne ? Une étude dédiée aux seules considérations « technico-techniques » de l'IA dans le secteur militaire serait vaine tant les perspectives d'évolution des systèmes d'algorithmes s'avèrent vertigineuses. Non, la réflexion que nous avons choisi de développer dans les prochaines pages vise plutôt à aborder les enjeux politiques, sociaux, organisationnels, stratégiques, juridiques et parfois éthiques d'un présent et d'un futur dans lesquels une part de plus en plus étendue de l'analyse stratégique et de la décision en matière politico-militaire sera imprégnée de traitements algorithmiques. Si l'accroissement des performances techniques des armements et des technologies avancées dépend essentiellement de variables financières – comme

ont pu nous le suggérer avec certitude certains représentants des milieux industriels de la défense –, les véritables incertitudes, elles, résident dans les variables humaines et organisationnelles. Autrement dit, comment nos institutions, nos bureaucraties, nos procédures de travail et modèles d'organisation parviendront-ils non seulement à intégrer mais aussi à absorber les transformations induites par l'expansion de l'intelligence artificielle ?

3. Définitions et incertitudes

Traiter de l'intelligence artificielle au sein d'une telle étude laisserait supposer que nous sommes en mesure de définir la notion même d'intelligence. Sans cela – serait-on tenté d'arguer –, quel intérêt pourrait-il y avoir à aborder la place de l'IA dans le champ de la défense ? Étudier la contribution avérée ou supposée de l'intelligence artificielle à nos systèmes militaires exige que nous fassions le deuil d'une définition de plusieurs concepts alors même que nous les manipulerons à foison. Ainsi, la notion même d'intelligence ne recevra-t-elle pas une définition claire. En vérité, cette question se situe à la croisée des champs d'investigation de la psychologie cognitive, de la philosophie et de la biologie. L'intelligence artificielle, ici entendue comme champ d'étude, avait pour objectif initial d'offrir une voie d'investigation nouvelle à la compréhension des fonctions cognitives humaines et, donc, de l'intelligence biologique chez l'homme. Cet objectif semble avoir été négligé avec le temps au gré des évolutions diverses de l'intelligence artificielle. Actuellement, face aux difficultés inhérentes à la définition de l'intelligence, on observe que nombre d'ouvrages et d'articles dédiés à l'IA ont préféré faire l'impasse sur la question de la définition. Il s'agit là, le plus souvent, d'un choix méthodologique destiné à avancer sur la question de la reproduction – ou de l'imitation – par la machine de ce que l'on pense être le mécanisme de l'intelligence. Dans son ouvrage consacré aux machines apprenantes, Yann Le Cun, directeur de la recherche au sein de FAIR (un département de Facebook dédié à la recherche sur l'IA), aborde la définition de l'IA sans même s'arrêter sur les éléments qui composent l'IA. En affirmant que l'intelligence artificielle consiste en la capacité d'une machine à reproduire des tâches généralement réalisées par les animaux ou les humains, Yann Le Cun évacue méthodologiquement la définition de l'intelligence.

Certes, sur le plan philosophique, une opposition semble s'être établie entre, d'une part, les experts convaincus que l'intelligence est d'abord une affaire de puissance de calcul et, d'autre part, les tenants d'une vision exclusivement biologique de l'intelligence. Il ne sera pas rare que nous croisions au cours de ces pages des considérations émises par quelques technologues et futurologues laissant penser que la principale différence entre l'homme et les « machines intelligentes » (expression à employer avec prudence) réside dans un différentiel de puissance de calcul. Les partisans de la *singularité technologique*[7] (ou plus simplement « *singularité* ») affirment que l'intelligence de la machine a d'ores et déjà dépassé les fonctions cognitives humaines. Les opposants à cette vision machiniste défendent la « supériorité » de l'homme sur la machine à l'aide d'arguments identiques en soutenant que la complexité et la puissance de calcul du cerveau humain ne pourra jamais être égalée par la machine.

[7] En mathématiques, une *singularité* peut être définie comme un point, une valeur ou une situation dans laquelle un objet mathématique ne peut être clairement défini ou subit une transition. Par extension à d'autres domaines, la notion de *singularité* a pu se rapporter à des objets très divers tout en signifiant l'indéfinition ou la phase de transition de l'objet considéré. Ainsi, en relativité générale, une singularité correspond-elle à une région de l'espace-temps au voisinage de laquelle certaines quantités décrivant le champ gravitationnel deviennent infinies quel que soit le système de coordonnées retenu, à l'exemple d'un trou noir dont le cœur incarnerait une singularité de cet ordre. La *singularité technologique*, quant à elle, s'éloigne quelque peu des mathématiques et de la physique et décrit tout à la fois une phase de transition et une situation nouvelle. L'hypothèse défendue par les partisans de la *singularité technologique* consiste à affirmer que l'explosion des capacités de stockage et de calcul des ordinateurs conduira à une transition entre machines intelligentes et machines conscientes. La *singularité technologique* désigne tout à la fois ce franchissement et l'existence même d'une intelligence artificielle consciente.

Cette vision des rapports entre l'homme et la machine est essentiellement quantitative. D'autres lui opposent une vision qualitative en soulignant que l'intelligence du vivant est fondamentalement supérieure à toute forme d'intelligence synthétique, même à supposer que cette dernière puisse être en mesure de dépasser, pour certaines tâches et opérations, les aptitudes cognitives humaines.

Ce débat de nature essentiellement philosophique sera abordé partiellement par notre étude là où les questionnements liés au statut de l'intelligence humaine et synthétique seront de nature à influer sur les interrogations de portée stratégique et géopolitique. Nous renverrons le lecteur intéressé spécifiquement par ces débats à des sources bibliographiques qui lui sont consacrées.

4. Organisation de l'étude

Le parcours réflexif que nous proposons dans le cadre de cette étude se déploiera comme suit. Le premier chapitre aura pour objectif de revenir sur la notion même d'intelligence artificielle et de comprendre l'origine du terme apparu au lendemain de la Seconde Guerre mondiale. Cette partie vise à prendre conscience de ce en quoi le champ d'études de l'IA, dans les années 1950, est l'héritage direct de l'entreprise sans précédent de recherche scientifique mise sur pied dans le contexte du renversement des puissances de l'Axe et de ce qui allait devenir la lutte entre les deux superpuissances de l'après-guerre. Ce chapitre propédeutique s'attardera sur les méandres de la recherche dans le domaine de l'IA et les soubresauts qu'elle a pu connaître par le passé, parfois déchirée entre plusieurs écoles. Nous clôturerons ce chapitre par une mise au clair des multiples notions qui entourent l'intelligence artificielle, tantôt prise en tant que discipline, tantôt en tant que système technique.

Le second chapitre s'attardera sur les perspectives d'applications militaires des systèmes d'intelligence artificielle, tels que nous les concevons aujourd'hui. Loin de constituer un catalogue des multiples affectations de l'IA pour les besoins du soldat et des états-majors, il s'agira surtout d'isoler en « grands ensembles » les domaines du combat dans lesquels l'IA pourrait apporter une plus-value pour la conduite des opérations militaires à moyen terme. Nous envisagerons donc ici essentiellement les bénéfices supposés, attendus, mais aussi réels de l'IA en matière de défense.

Un troisième chapitre sera quant à lui consacré aux scénarios stratégiques élaborés à partir de l'hypothèse de l'intégration de systèmes d'intelligence artificielle au sein des appareils militaires. Nous nous concentrerons, plus spécifiquement, sur deux interrogations : la première consistera à déterminer dans quelle mesure la perspective de l'adjonction de systèmes avancés d'IA serait de nature à précipiter le déclenchement d'un conflit au cours duquel des États auraient pour objectif d'empêcher un de leurs homologues de se doter d'une telle technologie. La seconde interrogation consistera à nous demander si l'aide à la décision qu'apporteraient des systèmes d'IA à la construction de la décision stratégique entraînerait une accentuation des tensions internationales et le déclenchement de guerres qui, en l'absence d'IA, auraient été évitées. Le lecteur percevra directement toute la difficulté de répondre à cette double interrogation de manière univoque et péremptoire. Aussi, notre démarche reposera-t-elle sur une exploration de plusieurs schémas prospectifs proposés par divers auteurs qui se sont penchés sur les mécanismes susceptibles de se déclencher dans les équilibres militaires et les relations géostratégiques du fait de l'expansion de l'IA au sein des édifices militaires des nations les plus avancées. Ce troisième chapitre nous permettra également d'aborder avec plus de profondeur l'épaisseur particulière du champ de la dissuasion nucléaire si celle-ci devait s'appuyer, de manière partielle ou complète, sur des dispositifs assistés par intelligence artificielle, qu'il s'agisse des mécanismes d'alerte rapide, de la détection de tirs, de l'analyse des données, des solutions sur le champ de bataille ou de la prise de décision. Comme nous aurons l'occasion de l'étudier, les implications de l'IA en matière de dissuasion nucléaire présentent de multiples facettes et les résultats des investigations restent toujours soumis à une part d'incertitude. Celle-ci s'explique très certainement par la nature spécifique des enjeux de la dissuasion nucléaire.

Un quatrième chapitre, que nous avons souhaité dissocier du précédent, même s'il comporte des similitudes au niveau des *approches*, portera sur les scénarios à dimension globale d'un futur dans lequel une intelligence forte, voire une *superintelligence*, émergerait. Baptisé, non sans une certaine ironie, « IApocalypses », ce chapitre évoquera l'hypothèse de « mondes » dans lesquels l'IA viendrait à s'établir comme un acteur à part entière de la régulation politique, sociale et économique. Au-delà des dystopies qu'ils véhiculent, les hypothèses à long terme qui y seront développées sont le produit de réflexions menées par d'éminents scientifiques qui, bien que n'étant pas tous impliqués au même degré dans le développement de l'intelligence artificielle, savent la place croissante qu'ont occupé les technologies numériques les plus avancées dans la conduite de leurs recherches. Ce quatrième chapitre s'attardera plus longuement sur le risque de « fracture herméneutique » que seraient susceptibles de générer les systèmes d'intelligence artificielle les plus avancés. L'extension exponentielle des secteurs d'activités conquis par l'IA conjuguée à l'accroissement des puissances de calcul et à la complexité croissante des réseaux neuronaux contribue à rendre le fonctionnement des systèmes d'IA hors de portée de l'intellect humain. Ce danger, que d'aucuns jugent imminent (à l'instar d'Elon Musk), conduit paradoxalement à des prises de position inattendue à travers lesquelles des experts appellent à une augmentation des capacités cognitives humaines à l'aide d'implants cérébraux. L'homme augmenté, mélange de carbone et de silicium, serait le salut de l'Humanité. D'autres experts encore alertent l'opinion sur les biais de l'IA, soit qu'ils sont inhérents à la machine même (aussi évoluée et performante soit-elle), soit qu'ils découlent de l'interaction altérée entre l'homme et la machine intelligente. Aurons-nous, demain, affaire à des sytèmes d'IA « volontairement » complaisants, soit qu'ils auront été programmés à cet effet, soit qu'ils auront appris à l'être en vue d'optimiser leurs rapports avec les humains ? Nous tâcherons de donner des pistes de réponse à ces quelques interrogations.

Dans le cadre d'un cinquième chapitre, nous inviterons le lecteur à se pencher sur les « géopolitiques » de l'IA. La maîtrise des technologies de l'intelligence artificielle ne constitue pas seulement une entreprise technoscientifique ; elle est aussi le lieu de luttes de pouvoir, le terrain de rivalités entre puissances établies et en devenir. Pour quelques acteurs, à l'instar des États-Unis, le leadership dans le secteur de l'IA vise le maintien de leur supériorité politique et militaire au cours du XXI^{e} siècle, à l'heure même où celle-ci pourrait être remise en question. La conquête de l'intelligence artificielle la plus avancée exige donc un contrôle absolu, sinon total, de l'ensemble des segments industriels de la filière. Qu'il s'agisse de l'obtention des matières premières (et de la garantie d'approvisionnement en terres rares), de la maîtrise en ingénierie ou du pouvoir d'attraction des meilleurs chercheurs et ingénieurs de la planète, le statut de leader dans le domaine de l'IA – que visent de nombreuses nations – impliquera des rivalités géopolitiques majeures. En effet, à côté des puissances technologiques « historiques », de nombreuses nations désireuses de s'affranchir de la grammaire occidentale des relations internationales voient dans la maîtrise de l'intelligence artificielle le tremplin idéal pour s'imposer comme puissance économique et politique globale. Pour des raisons spécifiques à chacune d'elles, la Chine et la Russie font partie de ces acteurs pour lesquels le contrôle de l'ensemble des éléments de la filière de la nouvelle industrie du numérique et des technologies qui lui sont liées n'est ni plus ni moins qu'un impératif de premier ordre. Preuve de la dimension géopolitique de l'IA : au sein même de l'Alliance atlantique, des réflexions se sont fait jour parmi les États membres sur le caractère stratégique de cette technologie et une certaine coordination de programmes est opérée au sein des agences dédiées à la recherche scientifique et technologique.

Dans un sixième et dernier chapitre, nous avons choisi de nous intéresser aux réflexions portées par quelques scientifiques, organisations non gouvernementales et internationales à propos des systèmes d'armes létaux autonomes (SALA). On pourra nous reprocher de greffer à cette étude une partie consacrée à une problématique qui, d'un point de vue technique, ne semble pas relever à proprement parler de l'intelligence artificielle. Si nous avons fait le choix d'aborder les débats entourant les SALA,

c'est en raison de la confusion qui semble s'installer entre le statut futur de ces armements (pour l'instant seulement redoutés, puisqu'inexistants) et l'IA. L'autonomie supposée de ces systèmes d'armes est souvent associée à l'idée d'une intelligence artificielle embarquée ou télépilotant ce type d'armement. Cet amalgame – que nous tenterons d'analyser – procède avant tout d'une manœuvre argumentaire de la part de certains observateurs visant à confondre dans un même ensemble, en raison des images d'Épinal qu'ils véhiculent, des systèmes d'armes dont la mise en œuvre n'engendre point de controverse sur le plan du droit international, d'une part, et des visions fantasmées de la guerre du futur, d'autre part. Bien que certains arguments avancés par les détracteurs des SALA peuvent être admis, il convient de procéder à quelques dissociations entre les notions débattues. Ce n'est qu'à travers un tel travail de catégorisation qu'une réflexion constructive pourra émerger quant au devenir de la guerre, du droit qui régit cette dernière et de nos organisations de défense.

Bien qu'aucune étude portant sur un objet technique aussi complexe et « maltraité » que l'IA ne puisse aisément se prêter à une forme de distanciation intellectuelle (qui serait pourtant salutaire et impérative), nous avons tenté de nous astreindre à une certaine objectivité. La présente étude ne constitue pas davantage un plaidoyer en faveur de l'IA qu'un réquisitoire à son encontre. L'Histoire regorge de révolutions techniques qui, réprimées dans les premiers temps, ont fini par s'inscrire dans l'ensemble des secteurs d'activités de la vie humaine et revêtir une normalité dans leur emploi. Notre propos s'efforcera d'éviter toute forme d'exclusive dans l'approche de l'intelligence artificielle. Nous réfutons ainsi une vision uniquement « essentialiste » de l'IA, de la même manière que nous nous attachons à éviter une approche purement « utilitariste » de celle-ci. En effet, bien qu'il soit difficile d'admettre qu'une technologie soit « mauvaise » en soi, on ne peut nier la spécificité de la trajectoire que peuvent présenter les évolutions de l'IA. Dans une acception « ellulienne » que nous qualifierons de « mesurée », il nous faut admettre que les développeurs de l'intelligence artificielle se veulent porteur d'un projet *rationnalisant*, tout aussi artificiel qu'il *artificialise* le monde, pour faire naître un environnement nouveau censé remplacer l'ancien. Il est troublant de constater que l'IA est l'archétype même d'un « système technicien » dénoncé par de nombreux philosophes qui voient dans les technologies issues des diverses révolutions industrielles un projet globalisant, fondé sur une interconnexion toujours plus poussée des artefacts et des secteurs dans lesquels ils sont présents. L'IA, en ce sens, est un projet global effaçant les différences entre les États qui en auront la maîtrise et accroissant la fracture avec les acteurs qui en seront dépossédés. Pourtant, on ne saurait nier que l'IA, comme toute technologie, est également le fruit de ses concepteurs et peut-être plus encore de ses utilisateurs. Les *Science & Technology Studies* nous enseignent, en effet, que les conditions de propagation et de mise en œuvre d'une technologie constituent également la résultante des utilisateurs de cette technologie. Comme nous aurons l'occasion de le voir au cours de ce chapitre, même dans sa version plus « primitive » ou « première », l'intelligence artificielle, qui n'était alors qu'automation, n'a jamais été abandonnée à elle-même, l'homme ayant toujours décidé de conserver l'ultime instance décisionnelle. Le but de notre recherche sera de déterminer si tel sera toujours le cas.

I. Intelligence artificielle : de quoi s'agit-il ?

Que recouvre, aujourd'hui, la notion d'intelligence artificielle ? À elle seule, cette question pourrait générer une multitude d'ouvrages scientifiques sans qu'elle ne reçoive pour autant une réponse satisfaisante, non pas dans sa rigueur, mais pour qui serait à la recherche d'une approche unitaire et univoque du terme. Aucun sujet technique, à dire vrai, ne saurait faire l'objet d'une définition holistique et réductrice. Ceci est d'autant plus vrai lorsque l'on parle d'intelligence artificielle. La particularité de l'IA est d'avoir été depuis plusieurs décennies une branche peu connue de la science qui s'est révélée surtout par son expression technologique ultérieure[8]. La qualification d'intelligence artificielle relève par ailleurs de la convention, qui elle-même dépend de l'époque de cette qualification.

1. D'abord, un domaine d'étude…

L'intelligence artificielle désigne, en premier lieu, une discipline scientifique née en 1956 au Dartmouth College de Hanover dans l'État du New Hampshire aux États-Unis. Le terme « intelligence artificielle » fut d'ailleurs inventé par John McCarthy qui, avec Marvin Minsky, Nathaniel Rochester et Claude Shannon, avaient jeté les bases d'une réflexion destinée à mieux comprendre toutes les formes d'intelligences (humaine, animale) en tentant de reproduire sur une machine ce que l'on pensait être ses mécanismes constitutifs. En d'autres termes, c'est pour mieux comprendre l'intelligence humaine (l'apprentissage, les facultés de perception, le calcul, la mémorisation, la création artistique) que naquit l'intelligence artificielle (entendue comme domaine d'étude). Ainsi, au cours des soixante années qui ont suivi l'apparition de l'IA comme domaine d'études, de nombreuses fonctions cognitives humaines ont-elles été simulées informatiquement afin de mieux comprendre les processus qui les sous-tendent. Ces recherches menées dans le cadre de l'IA ont parfois emprunté des approches spécifiques et ont connu nombre de réorientations, compte tenu des impasses parfois rencontrées lors des expérimentations opérées. Il n'en demeure pas moins que ces diverses simulations ont permis de mieux comprendre toute la complexité des opérations qui se situent à la base de nos mécanismes cognitifs les plus courants et ont permis d'avancer sur le développement de nombreuses technologies aujourd'hui répandues dans notre quotidien.

« Les machines peuvent-elles penser ? » Telle fut la question posée en son temps par l'un des pionniers de l'intelligence artificielle, Alan Turing, dans le cadre d'un article scientifique qui fut très rapidement considéré comme le fondement des recherches dans le domaine de l'IA. Rédigé en 1950, cet article intitulé « Computing Machinery and Intelligence » exposait les premières tentatives de Turing pour mettre au point une machine susceptible de rivaliser avec le raisonnement humain[9]. Alan Turing fut, comme on le sait, le concepteur du décodeur de l'instrument de cryptographie allemand Enigma durant la Seconde Guerre mondiale. C'est sur la base des travaux qu'il conduisit durant cette période qu'il poussa par la suite ses réflexions sur la possibilité de rendre une machine capable d'un raisonnement proche de celui de l'humain. Le test dit « de Turing » visait expressément à la mise au point d'une machine susceptible, de par les réponses qu'elle émettait lors d'interactions élémentaires de langage, de conduire un agent humain à conclure à la nature humaine de son interlocuteur. Dans le cours des années 1950, les premières applications apparaissent. Arthur Samuel développe la

[8] André-Yves PORTNOFF, Jean-François SOUPIZET, « Intelligence artificielle : opportunités et risques », *Futuribles*, volume 5, numéro 426, 2018, p. 6.

[9] N. J. NILSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*? Cambridge, Cambridge University Press, 2009.

première machine s'appuyant sur un programme d'apprentissage automatique capable de battre des joueurs au jeu de dames.

Le contexte des années 1950 s'avère, par ailleurs, des plus propices à la recherche dans le domaine de l'IA. L'effort scientifique engagé dans le cadre de la Seconde Guerre mondiale a accouché d'une multitude d'avancées tant dans le domaine des sciences dures que dans celui des sciences humaines. Ce sont surtout les intersections nombreuses entre ces deux ensembles qui ont abouti à une large variété d'innovations à travers les champs disciplinaires. Les disciplines de l'IA et de la cybernétique, issues des travaux scientifiques qui ont accompagné le projet, se sont mutuellement renforcées pour constituer un berceau de recherche commun, porté du reste par les mêmes figures scientifiques. Il faut, en effet, comprendre le projet philosophique dans lequel s'inscrit dès son origine la discipline de l'IA ; la compréhension de ce projet ne peut être dissociée du champ de la cybernétique. C'est Norbert Wiener, professeur d'ingénierie et de mathématique au Massachussetts Institute of Technology (MIT), qui le premier envisagea la possibilité d'un croisement entre l'ingénierie et la biologie. Norbert Wiener avait un attrait marqué pour ce qu'il appelait les « zones frontières » de la science, autrement dit les points de rencontre entre des domaines de recherche en apparence distants. De cette rencontre entre l'ingénierie et la biologie naquit la *cybernétique* et son concept de rétroaction. La rétroaction est à l'origine un mécanisme issu de la biologie. Celui-ci permet l'autorégulation des êtres vivants et leur adaptation à l'environnement dans lequel ils évoluent. La cybernétique, aussi appelée science du contrôle (ou de la gouvernance[10]), s'est inspirée de la notion de rétroaction pour l'envisager dans le cadre des « transactions sociales ». Durant la Seconde Guerre mondiale, Norbert Wiener[11] et ses confrères avaient travaillé à la conception de procédures de rétroaction pour la stabilisation des batteries antiaériennes contrôlées par radar. C'est dans ce cadre que furent développées les premières hypothèses de la cybernétique. Bien que le concept de rétroaction soit aujourd'hui devenu particulièrement répandu et semble évident, tel n'était pas nécessairement le cas au lendemain de la guerre. La cybernétique introduisait ainsi une véritable révolution dans les conceptions de l'époque largement inspirées de la psychanalyse freudienne. Selon la vision freudienne, l'esprit ne faisait essentiellement que manipuler des énergies biologiques qu'il était dangereux de réfréner sous peine de les voir ressurgir sous des formes plus pernicieuses. La cybernétique, pour sa part, affirmait que le cerveau humain procédait avant tout du traitement d'informations issues de l'environnement. Les mécanismes cognitifs par l'intermédiaire desquels les informations étaient traitées constituaient une sorte de « boîte noire » et tout l'enjeu de la cybernétique était de comprendre le fonctionnement de cette boîte noire.

En dépit de l'épouvante que produit sur nombre d'esprits l'irruption de l'arme atomique mais aussi de l'impact imprévisible de cette arme nouvelle sur les équilibres militaires venant à se reformer, l'entreprise qui la sous-tend a permis le triomphe de la *technoscience*. Le projet cybernétique visant le développement d'une *machine intelligente* ou encore d'une *machine à penser* avait précisément pour but, comme l'évoque Céline Lafontaine, de « purifier la science du péché nucléaire ».[12] Tout comme le mouvement cybernétique, le champ disciplinaire naissant de l'IA s'inscrit dans un héritage militaire. En prônant la conception à terme d'une *machine intelligente*, les cybernéticiens envisageaient un projet de contrôle technocratique de la société dans le climat de la guerre froide. L'ambition était de

[10] Le verbe latin *gubernare* vient d'ailleurs du verbe grec *kubernein*.

[11] En dépit de l'apport de la cybernétique à la discipline que constitue l'IA, Norbert Wiener n'attachait pas une grande importance aux ordinateurs et à leurs possibilités. Son approche de la cybernétique s'inscrivait d'abord dans une logique de gouvernance dans laquelle la place des machines n'était pas jugée primordiale ou essentielle. Cf. Michel CREVIER, *À la recherche de l'intelligence artificielle*, Paris, Flammarion, coll. Champs, 1997, p. 45.

[12] Céline LAFONTAINE, *L'empire cybernétique : de la machine à penser à la pensée machine*, Paris, Seuil, coll. Essai, 2004, p. 48. Voir aussi Michel MORANGE, *Histoire de la biologie moléculaire*, Paris, La Découverte, 1994.

contourner l'irrationalité de l'Homme considérée comme la source des deux cataclysmes guerriers qui avaient secoué le XX^{e} siècle. Aussi, la création d'une machine pleinement rationnelle portait avec elle l'espoir d'une gestion plus juste et efficace de la société[13]. Cet objectif – la mise au point d'une machine à gouverner – était pour les membres de l'école cybernétique le projet ultime de leur mouvement. Pourtant, c'est le militaire qui représentera le véritable premier « client » de la cybernétique. Dès 1949 et le premier essai de bombe atomique de l'Union soviétique, les forces armées des États-Unis recoururent aux travaux des cybernéticiens en finançant abondamment leurs recherches au niveau de leur composante « informatique ». Il s'agissait pour les États-Unis de permettre la conception d'un système de surveillance pour l'organisation de la riposte en cas d'attaque nucléaire soviétique. Le projet SAGE (Semi-Automatic Ground Environment) fut entamé dès l'année 1950 et faisait de la cybernétique une discipline au service de la guerre. SAGE constitua le premier dispositif non humain utilisé pour l'analyse de l'information et l'orientation des décisions en temps réel ou quasi réel. Comme le faisait remarquer Philippe Breton, pour la toute première fois, « non seulement la machine remplaçait l'homme, mais elle agissait dans un univers temporel si rapide que l'homme n'y avait accès qu'après coup. »[14] Cette préoccupation se situe toujours au cœur des interrogations contemporaines relatives à la place de l'IA au sein des systèmes de gouvernance, quels que soient les secteurs au sein desquels ceux-ci sont déployés.

À partir des années 1960, deux écoles semblent émerger au sein de la discipline de l'IA. La première, dite « connexionniste », propose de s'inspirer de la nature pour parvenir à reproduire le fonctionnement du cerveau humain, l'idée sous-jacente étant que le système cognitif humain peut être reproduit au sein d'une machine « à penser ». L'approche connexionniste s'inspire en cela de la cybernétique. Les tenants de la seconde école, dite « symbolique », estiment que le cerveau humain s'avère tout à la fois plus complexe qu'on ne le pense et, surtout, mal adapté à une reproduction sur la machine. Ses membres s'orientent vers le développement de « systèmes experts » dont la mise au point s'appuie sur la maîtrise de prédicats logiques. Selon l'école symbolique, toute résolution d'un problème par une machine suppose, au préalable, la décomposition d'un raisonnement afin de codifier celui-ci en une série d'actions logiques que la machine pourra produire[15].

À la fin des années 1960, les premières désillusions se font jour. C'est le début de ce que l'on appellera bientôt « les hivers de l'intelligence artificielle ». L'école symbolique connaît des revers notoires en se heurtant à d'importantes difficultés dans la transposition du langage humain en codes informatiques. À cette même époque, le Department of Defense, qui avait déjà investi des sommes importantes dans la recherche sur l'IA, manifestait ses premiers signes d'impatience. Le monde militaire exprima de sérieux doutes sur la capacité des trois principaux centres investis dans cette discipline (le Massachussetts Institute of Technology, Carnegie Mellon et Stanford) à respecter leurs engagements[16]. Un autre problème de taille, autre que la logique adoptée pour les processus, vit le jour : la limitation des puissances de calcul des ordinateurs[17]. C'est finalement l'ensemble de la discipline naissante que constitue l'IA qui se voit affectée par les échecs rencontrés durant cette période.

[13] Céline LAFONTAINE, *op. cit.*, 2004, p. 50.
[14] Philippe BRETON, *Une histoire de l'informatique*, Paris, Seuil, 1990, p. 129.
[15] Daniel CREVIER, *À la recherche de l'intelligence artificielle*, Paris, Flammarion, coll. « Champs », 1997, pp. 134 et ss.
[16] Au cœur de la guerre froide, le DoD avait procédé à des dépenses considérables pour le développement de systèmes de décryptage et de traduction automatiques du langage. Au cours de la Seconde Guerre mondiale, le déchiffrement des codes secrets et messages allemands par la machine conçue par Alan Turing avait permis d'espérer en la capacité future des ordinateurs à traduire tous types de code et de langage. La CIA s'était montrée très tôt intéressée par la conception d'une machine de traduction automatique des publications soviétiques.
[17] Hans MORAVEC, *The Role of Raw Power in Intelligence,* Stanford, Stanford Libraries, 12 mai 1976.

Une renaissance relative survient au début des années 1980 avec les premiers systèmes experts[18]. De nouveaux langages informatiques (Prolog, LISP) suscitent des espoirs nouveaux. La DARPA (Defense Advanced Research Project Agency), aux États-Unis, multiplie à nouveau les programmes et les investissements. Cependant, toujours articulés autour de systèmes experts particulièrement fastidieux et lourds dans la mise en œuvre, les progrès en matière d'IA tardent à venir. L'informatique est à l'heure des ordinateurs personnels que développent des sociétés privées telles que Apple ou IBM. Les années 1990 n'offriront pas davantage de résultats, même si quelques événements semblent donner lieu à des pistes prometteuses. Ainsi en fut-il de la victoire de l'ordinateur *Deep Blue* sur le champion russe des échecs Gary Kasparov. En vérité, ce haut fait dans le domaine de l'IA marque surtout la victoire de l'école « symbolique » sur l'école connectiviste. La victoire de *Deep Blue* résida dans la capacité de l'ordinateur à traiter, face à son adversaire humain, plus de 200 millions de combinaisons de jeu possibles par seconde[19]. En dépit du choc que la victoire de *Deep Blue* put créer dans l'opinion, le secteur de l'IA ne parvint pas à rebondir sur cet événement. Les années 1990 restent marquées, certes, par l'exposition au grand jour de la compétition homme/machine naissante, sans toutefois connaître des percées substantielles au niveau qualitatif. Pas au point de faire naître une réelle rupture.

Les années 2010 se révéleront, quant à elle, le théâtre de bonds technologiques majeurs qui porteront la recherche et le développement dans le domaine de l'IA vers des horizons nouveaux. Il faut dire que le renouvellement de la discipline a bénéficié de la contribution du développement de l'internet et des autres TIC. Des champs d'application nouveaux ont ainsi émergé dans des domaines tels que le raisonnement symbolique, la modélisation, la recherche des données, les modes de représentation des connaissances ou encore le traitement du langage naturel. Aujourd'hui, des technologies tels que les filtres anti-spam de nos boîtes de messagerie électronique, les suggestions personnalisées de films et de musiques sont issues de la recherche en IA.

Mais là où l'apport de la numérisation a été le plus notable concerne la multiplication des données disponibles ; ces mêmes données qui nourrissent les systèmes d'IA et permettent leur entraînement. La profusion des données, rendue possible par des politiques de données ouvertes (*open data*), ainsi que les progrès techniques dans le champ des mégadonnées (*big data*) et des puissances de calcul autorisent le traitement de stocks immenses de bases d'informations dans des formats et registres variés.

Enfin, les développements récents intervenus dans la discipline de l'IA découlent aussi de l'ouverture de plusieurs technologies autrefois soumises à des restrictions propriétaires. Des acteurs tels que IBM, Google, Facebook ont choisi d'ouvrir à la communauté des chercheurs spécialisés dans l'IA nombre de technologies qu'ils avaient développées en interne. Cette politique a permis l'apport de nouveaux enrichissements et de nouvelles fonctionnalités.

Comment définir l'IA ? On s'en doutera, une telle question ne peut donner lieu à une réponse univoque. Pour l'informaticien Yann Le Cun, directeur d'études chez Facebook, l'IA « doit permettre de faire faire aux machines des activités qu'on attribue généralement aux animaux et aux humains »[20]. Une telle définition n'est pas sans poser quelques difficultés. Tout d'abord, elle ne correspond pas au critère d'une définition puisqu'elle n'évoque pas « l'essence » même de l'IA, mais uniquement les perspectives de transfert de tâches des animaux/humains aux machines. Ce faisant, elle assimile d'une part l'homme à l'animal, mais plus encore pose le vivant et l'IA comme équivalents. Cette approche

[18] Olivier KEMPF, « IA, explicabilité et défense », *Revue Défense Nationale,* mai 2019.
[19] Jean-Gabriel GANASCIA, *Le mythe de la Singularité*, Paris, Seuil, coll. Science Ouverte, 2017.
[20] Libération/France Inter, *Comment l'intelligence artificielle va changer vos vies*, Paris, Libération/France Inter, décembre 2017, p. 11.

laisse à penser que l'homme et la machine sont dès lors interchangeables à souhait, ce qui risque de susciter des contestations sociales.

Pour Cédric Villani, rapporteur d'une mission parlementaire consacrée au développement d'une stratégie française en matière d'IA, cette dernière doit essentiellement être abordée comme un « projet ». Plus spécifiquement, il s'agit du projet visant à comprendre la cognition humaine et à tenter de reproduire celle-ci de manière artificielle. En d'autres termes, dans la définition citée plus haut, l'IA n'est pas tant définie par rapport à ce qu'elle est que par rapport à ce à quoi elle pourrait aboutir sur le plan scientifique. En vérité, cette définition pose un certain nombre de difficultés au regard de l'histoire de la discipline. Si les premières tentatives de développement d'une intelligence de type « machine » eurent pour ambition de reproduire artificiellement (au sein d'un artefact, donc) les mécanismes cognitifs humains, il apparut progressivement qu'une telle projection se heurtait à nombre de problèmes, dont celui de la puissance des ordinateurs et de la disponibilité des données. La piste vers le développement d'un système d'IA fort et capable de reproduire le cerveau humain fut donc considérablement reconsidérée.

On le voit, toute tentative de définition de l'intelligence artificielle offre une pluralité de lectures. Peut-être la difficulté ne vient-elle pas tant du signifié que du signifiant. Pour être plus clair, ce n'est peut-être pas la réalité de l'IA qui porte à confusion mais bien l'expression « intelligence artificielle » en elle-même. C'est là, en tous cas, le point de vue mis en avant par Patrick Bezombes, Directeur adjoint du Centre interarmées de concepts, de doctrines et d'expérimentation (CICDE). Selon ce dernier,

« les débats sociétaux et l'agitation médiatique auxquels nous assistons sur l'IA n'auraient probablement pas eu lieu si nous avions gardé la terminologie scientifique et technique initiale, certes peu attrayante, à savoir « traitement du signal », « traitement des données », « algorithmie » et « automatisme »[21].

2. L'IA en tant que système

Évoquer l'IA en tant que tel ne rend qu'imparfaitement compte de la complexité du concept. Sans doute est-il préférable d'envisager l'IA comme un système. C'est en ce sens que la Commission européenne a récemment défini l'IA[22] :

> « *L'intelligence artificielle (IA) désigne les systèmes qui font preuve d'un comportement intelligent en analysant leur environnement et en prenant des mesures – avec un certain degré d'autonomie – pour atteindre des objectifs spécifiques.*
>
> *Les systèmes dotés d'IA peuvent être purement logiciels, agissant dans le monde virtuel (assistants vocaux, logiciels d'analyse d'images, moteurs de recherche ou systèmes de reconnaissance vocale et faciale, par exemple) mais l'IA peut aussi être intégrée dans des dispositifs matériels (robots évolués, voitures autonomes, drones ou applications de l'internet des objets, par exemple).* »

[21] Patrick BEZOMBES, « Intelligence artificielle et robots militaires », *Défense & Sécurité Internationale Hors Série*, numéro 65, avril-mai 2019, p. 12.

[22] COM(2018) 795 final, Communication de la Commission au Parlement européen, au Conseil européen, au Conseil, au Comité économique et social européen et au Comité des régions, *Un plan coordonnée dans le domaine de l'intelligence artificielle*, https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52018DC0795&from=DA.

D'une façon générale, les spécialistes de l'IA préfèrent employer la notion de « systèmes d'IA ». La mise en œuvre d'une IA suppose l'intervention d'une variété de dispositifs et de processus :

1. des *senseurs,* dont le rôle est de collecter les données issues de l'environnement dans lequel opère le système d'IA. Les types de senseurs dépendent directement du type de mission confié à l'IA par le concepteur ;
2. des *processeurs,* dont le rôle est d'interpréter les données collectées et de définir le mode d'action le plus adapté à la tâche confiée ;
3. une *boucle de rétroaction,* permettant au système d'IA de s'adapter au nouvel environnement issu de sa propre intervention.

À cette vision schématique et volontairement « générique » de l'IA, il faut ajouter les ruptures technologiques qui permettent à ce système de fonctionner. En réalité, ce sont les progrès intervenus dans quatre branches fondamentales des technologies de traitement de l'information qui ont permis, à partir de 2010, à l'IA d'aboutir à des résultats probants. Quelles sont ces technologies ?

Il y a tout d'abord les communications multipliant les réseaux à l'échelle planétaire. Aujourd'hui, l'ensemble de la planète (ou une immense majorité de sa population) est interconnectée par une toile informationnelle combinant tous les types de médias. Les réseaux de communication numériques mettent désormais en rapport la majorité des habitants de la planète et la presque totalité des dispositifs électroniques en transportant en temps réel voix, musique, images fixes ou mobiles, documents, etc. Cet ensemble est souvent désigné comme un « unimédia » puisqu'il n'existe plus de différence réelle de liaison ou de traitement entre les différentes formes de transfert de l'information. En 2016, à l'échelle du monde, on dénombrait 7.509 milliards d'abonnements. Quant à l'utilisation de l'internet, celle-ci est passé de 0,05 % de la population mondiale en 1989 (année de l'ouverture du web aux particuliers) à 45,8 % des habitants de la Terre. Les réseaux sociaux ont connu une évolution exponentielle similaire.

Vient ensuite la puissance de calcul. Celle-ci est actuellement au cœur d'une course entre les principales puissances technologiques et scientifiques de la planète, les États-Unis et la Chine en tête. L'intégration des circuits a autorisé la mesure des puissances de calcul en FLOPS (*Floating-point operations per second*). La capacité de calcul a longtemps suivi la loi de Moore qui, en 1965, annonçait le doublement tous les deux ans du nombre de transistors des microprocesseurs. Cette loi de Moore – loi essentiellement auto-réalisatrice et réduite à son expression la plus caricaturale –, a longtemps fait figure de référence en matière de prévision de l'accroissement de la puissance de calcul des ordinateurs. Elle rencontre cependant, aujourd'hui, des limites physiques que doivent permettre de dépasser les composés nanotechnologiques et les perspectives d'innovation dans le domaine du calcul quantique. Depuis 2008 et la mise sur le marché du superordinateur *Roadrunner*, la capacité maximale de calcul atteinte par l'ordinateur a franchi le million de milliards d'opérations par seconde (capacité exprimée en pétaFLOPS. La Chine et les États-Unis prétendent chacun être en mesure, d'ici 2021, de concevoir des supercalculateurs atteignant un milliard de milliards d'opérations par seconde (exaFLOPS). Pour l'heure, ce sont bel et bien les États-Unis qui remportent la course avec le *Behold Summit* et ses 200 pétaFLOPS, conçu par IBM et Nvidia, et basé au Laboratoire national d'Oak Ridge. Le *Taihulight* du National Supercomputing Center chinois occupe la seconde place en atteignant les 100 pétaFLOPS. La troisième position est celle occupée par le *Tihanhe-2*, également chinois, avec 34 pétaFLOPS. Tout récemment, l'entreprise Google a affirmé être parvenue à mettre en œuvre le premier calcul de type quantique, lui permettant de résoudre en seulement trois minutes un problème qui aurait pris quelque 10.000 ans au supercalculateur américain *Behold Summit*. Si cette nouvelle venait à se confirmer, elle impliquerait une transformation vertigineuse, inédite dans l'histoire de l'Humanité, dont les effets se feraient sentir dans l'ensemble des secteurs d'activité humaine. Une telle conjoncture est encore difficilement projetable pour l'esprit humain. Pour le technologue et

futurologue Raymond Kurzweil, la capacité de calcul quantique, une fois maîtrisée et généralisable à la résolution de plusieurs problèmes simultanés (Google affirme n'être actuellement parvenu qu'à concentrer son calculateur quantique sur la résolution d'un seul problème spécifique), devrait conduire à l'émergence d'une véritable *singularité* où la machine égalerait l'homme en intelligence. Depuis cette annonce, les experts prévoient une augmentation des capacités de calcul des ordinateurs quantiques à un taux exponentiel double (une fonction exponentielle dont l'exposant est lui-même une fonction exponentielle). À titre de comparaison, la loi de Moore, qui porte sur les processeurs classiques, se base sur une croissance exponentielle simple.

Un système d'IA apprenant ne pourrait être opérant sans disposer de données lui permettant d'alimenter ses bases de travail nécessaires aux opérations de corrélation, alimentation rendue possible en ce XXI^{e} siècle par la généralisation de l'outil informatique à travers une grande partie du monde, les modes de fonctionnement des infrastructures publiques et privées, nos multiples activités individuelles, collectives, notre (omni)présence sur l'internet à travers les réseaux sociaux, la collecte permanente de données privées (localisation GPS, temps de navigation sur la toile, photos, vidéos, publications, etc.). D'ici 2025, il est prévu que le volume planétaire de données transmises et stockées atteigne les 175 zettaoctets[23], soit 175 milliards de téraoctets. En guise d'illustration, un tel montant de données nécessiterait 12,5 milliards de disques durs de 14 téraoctets chacun. Le nombre de piles de disques Blu-ray qui pourraient stocker une telle quantité de données permettrait de couvrir 23 fois la distance Terre-Lune. Ce que l'on désigne également par le *big data* doit, pour pouvoir alimenter les systèmes d'IA, combiner trois caractéristiques résumées par les trois « V ». Il s'agit, tout d'abord, du volume. Nous venons à l'instant de l'évoquer : c'est la disponibilité d'une multitude de données, perpétuellement mises à jour, qui s'avère fondamentale pour l'éducation d'une IA. Ensuite, pour parler de *big data*, il est essentiel d'avoir affaire à des données variées. Des données sont considérées comme variées dès lors que leur provenance associe une large diversité de sources, parmi lesquelles figurent les objets intelligents, les capteurs, l'internet ou encore les technologies de collaboration sociale. La particularité de ces données est d'inclure, outre les données relationnelles traditionnelles, des données brutes (qu'elles soient non structurées ou partiellement organisées) provenant de pages web, de fichiers en ligne, d'index de recherche, de médias sociaux, de courriers électroniques, de tweets, de blogs, de vidéos, etc. Il est important de souligner que la mise à disposition des données n'est pas une valeur en soi. C'est leur structuration qui est valorisée. Or plus de 85 % des données récoltées auprès des utilisateurs à travers le monde sont tout simplement « non structurées ». Cela signifie que, pour tirer parti – et donc valoriser – des données du *big data*, un travail considérable d'analyse et de tri des données doit être opéré par les organisations. Le véritable enjeu du *big data* se situe précisément dans le traitement d'une multitude de bases de données afin de procéder à un maximum de recoupements entre les sources. Enfin, la vitesse de récolte, de traitement et de valorisation s'avère un ingrédient capital du *big data*. L'avènement de l'internet des objets exigera de disposer de systèmes capables de récolter et de trier en temps réel ou quasi réel les données infinies en provenance de tous types de terminaux et supports. Cette vitesse de traitement constitue un atout fondamental car elle permettra aux IA de disposer d'informations à jour dans le cadre de l'auto-apprentissage.

Le *big data* n'a de valeur qu'au travers d'algorithmes apprenants. C'est dans ce secteur que les révolutions technologiques les plus notables ont vu le jour durant ces vingt dernières années. Les choses ont réellement commencé à changer à partir de 2005, grâce à l'apprentissage automatique et à l'apprentissage profond (*deep learning*). L'apprentissage profond revisite, en quelque sorte, le « connexionnisme » des années 1960 et s'inspire grandement des récentes neurosciences (cf. l'actuelle vague des technologies NBIC). Les algorithmes permettent le traitement et l'interprétation

[23] Soit 10^{21} octets.

des données en recherchant au travers de celles-ci des corrélations ou régularités. À cela s'ajoute l'assistance de nouvelles techniques statistiques et de calculs de probabilités. Cependant, la nouveauté la plus marquante dans le domaine des algorithmes a vu le jour tout dernièrement avec l'émergence des modalités d'apprentissage profond (*deep learning*). Depuis les années 1980, on connaissait surtout les systèmes apprenants (*machine learning*). L'apprentissage profond, quant à lui, repose sur des réseaux de neurones, autrement dit des programmes informatiques structurés en couches reliées par des communications synaptiques dont l'architecture s'inspire du cerveau humain. De tels réseaux de neurones ont la faculté d'apprendre par eux-mêmes à partir d'une quantité gigantesque d'exemples. Ce sont ces nouvelles technologies algorithmiques qui permettent aujourd'hui à une machine de procéder à la reconnaissance d'images et de sons. Ce sont elles encore qui permettent aux machines d'interagir de manière naturelle avec l'homme.

a) Apprentissage profond et apprentissage automatique

L'apprentissage profond désigne un mode de traitement effectué par un grand nombre de neurones artificiels imitant de façon très simplifiée le cerveau biologique. Les interactions opérées permettent au système d'apprendre progressivement à partir d'images, de textes et d'autres types de données croisées. Cet apprentissage repose sur ni plus ni moins que des principes mathématiques généraux. Le résultat du processus d'apprentissage peut prendre plusieurs formes : une représentation (exemple : cette image comporte des éléments différents), une décision (exemple : cette image représente un chat) ou une transformation (exemple : traduction d'un texte d'une langue à une autre). L'apprentissage profond a donné une impulsion nouvelle à la recherche dans le domaine de l'intelligence artificielle, puisqu'elle a permis de ranimer les recherches dans le domaine de la vision par ordinateur, la reconnaissance automatique de la parole ou encore la robotique. C'est en 2012 que les premières applications issues de l'apprentissage profond voient le jour, notamment dans le domaine de la reconnaissance de la parole (application Siri sur l'iPhone de la société Apple). L'année 2012 marque également le point de départ – et ce n'est pas un hasard – des techniques d'IA basées sur les réseaux de neurones. En vérité, l'apprentissage profond constitue un mode de développement de l'IA relativement ancien : c'est en effet dans les années 1990 que cette technique a été imaginée pour la première fois. Elle s'était même avérée assez prometteuse, d'un point de vue théorique. Cependant, la poursuite des recherches dans le domaine de l'apprentissage profond butait contre les limites de la technologie de l'époque, principalement les capacités de traitement restreintes des processeurs[24]. L'arrivée sur le marché des processeurs graphiques de type GPU ont véritablement révolutionné les travaux et les réalisations dans le domaine de l'IA. Très rapidement, la communauté scientifique se rendit compte de la portée de l'apprentissage profond combiné au processeurs graphiques. En ce sens, il peut être dit que c'est principalement la reconnaissance d'images qui permit au champ disciplinaire de l'IA de réaliser un véritable bond en avant. Par la suite, de nouveaux logiciels assimilés ont étoffé la gamme des applications issues de l'apprentissage profond en intégrant, notamment, le moteur de recherche Google. Les grands groupes du numérique et de l'internet comprennent toute la portée révolutionnaire de l'apprentissage profond : Google s'adjoint ainsi en 2013 les services de Geoffrey Hinton pour son projet BRAIN tandis que Facebook engage cette même année le Français Yann Le Cunn pour la mise en place du laboratoire FAIR. C'est avec le recrutement de ces deux scientifiques que tout, au niveau de l'IA, a redémarré…

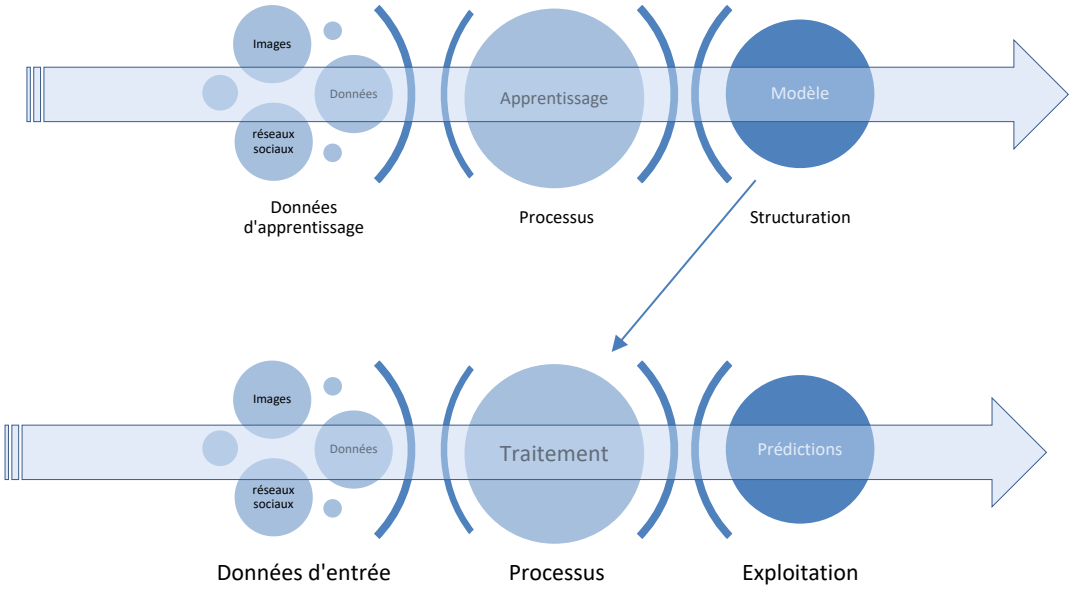
Dans le cadre de l'apprentissage machine, les connexions entre neurones constituant un réseau artificiel donnent lieu à des inférences complexes. Ces réseaux de neurones artificiels sont alors entraînés à l'aide d'un très grand nombre d'images (ou d'exemples) à établir des interconnexions entre

[24] Enki BILAL et al., *Intelligence artificielle. Enquête sur ces technologies qui changent nos vies*, Paris, Flammarion, coll. Champs/Actuel, 2017, p. 15.

divers éléments. En matière de reconnaissance faciale, par exemple, le réseau détermine lui-même les paramètres et critères devant lui permettre de reconnaître la présence d'un visage dans une image et, éventuellement, d'identifier la personne à laquelle appartient ce visage. Toutefois, si ce type d'apprentissage peut se révéler des plus intuitifs pour l'être humain (une personne parviendra à reconnaître un chat après avoir vu seulement une image de chat), l'intelligence artificielle exigera quant à elle un nombre aussi grand que possible d'illustrations lui permettant d'aboutir aux mêmes concordances. Ceci explique pourquoi il est souvent affirmé, à raison, que l'intelligence artificielle est moins une question de « programmation » que d'« éducation » de la machine. Pour l'IA, la détermination exacte des différents critères devant être pris en considération pour la reconnaissance d'une image ou d'un texte est fondamentale. Ce travail s'opère au travers de plusieurs couches de traitement. Une première couche consistera à repérer des pixels, une seconde couche d'analyse permettra d'identifier des formes géométriques susceptibles d'indiquer la présence d'un visage et une troisième couche de traitement détaillera le positionnement de la bouche, des yeux, des lèvres, etc. C'est au terme de ce travail et d'une comparaison des éléments issus de multiples bases de données que l'IA parviendra éventuellement à identifier socialement la personne dont le visage lui a été soumis.

L'apprentissage profond constitue un véritable sursaut de l'IA en tant que champ disciplinaire. Pendant de nombreuses années, la recherche dans le domaine de l'intelligence artificielle avait buté sur la manière de permettre à la machine de réaliser des tâches qui étaient évidentes pour l'homme. Cette difficulté découlait du fait que l'essentiel de la connaissance que l'homme a du monde autour de lui ne se présente pas sous la forme d'un ensemble de tâches explicites, processus qui est indispensable pour la programmation d'une machine. Aujourd'hui, l'apprentissage profond permet à la machine d'apprendre selon une méthodologie nouvelle qui se distingue des procédés antérieurs. Cet apprentissage profond s'inscrit par ailleurs dans le cadre plus général de l'apprentissage *automatique*. Il s'agit d'un ensemble de principes généraux permettant aux systèmes d'IA d'utiliser les données disponibles pour apprendre à « bien décider », à acquérir de bonnes connaissances et, au final, à rechercher des données nouvelles pour apprendre encore mieux et affiner leurs protocoles. L'apprentissage automatique, autrement dit la capacité pour une machine d'apprendre à apprendre, mobilise des disciplines aussi diverses que la biologie ou encore les sciences cognitives. Donner les moyens à une machine de prendre la bonne décision ne dépend pas seulement du nombre de bases de données à la disposition du système. Encore faut-il comprendre ce qu'est une « bonne décision ». Dans le règne animal, une bonne décision est peut-être et avant tout une décision permettant de garantir la survie d'un individu ou de son espèce (si l'on suppose l'existence d'une prise de conscience de groupe d'appartenance). Pour l'homme, la définition d'une bonne décision est autrement plus complexe tant elle convoque des variables sociales, culturelles, politiques et économiques susceptibles, du reste, d'entrer en contradiction entre elles. Surtout, l'apprentissage automatique se heurte à une limite qui semble ne pas concerner le cerveau humain. Notre cerveau est doté d'une aptitude à incorporer des algorithmes généraux lui permettant d'apprendre une multitude de tâches auxquelles l'évolution n'avait pas préparé ses prédécesseurs. Ainsi, le cerveau humain a-t-il été capable de bâtir des ponts, des routes, à jouer aux échecs sans qu'une intelligence tierce ne vienne lui inculquer de telles compétences. À l'inverse, dans le cadre de l'apprentissage automatique, l'existence d'un *algorithme universel* fait défaut. Pour chaque type de problème, un algorithme spécifique sera nécessaire. Bien sûr, on peut supposer que certaines catégories de tâches présentant des similitudes (dans certaines limites) puissent être prises en charge et résolues par un même algorithme ou un algorithme similaire. Toutefois, l'idée qu'un algorithme universel puisse suffire à résoudre n'importe quel problème relevant d'une multitude de champs doit être écartée. En d'autres termes, si tous les problèmes sont équiprobables, il n'existe pas d'algorithme universel, c'est-à-dire meilleur que tous les

autres sur l'ensemble des problèmes. Cette limitation constitue une difficulté majeure pour les tenants de l'intelligence artificielle générale ou superintelligence[25].



b) Recherche fondamentale et recherche appliquée

Sérendipité : la conjonction du hasard heureux qui permet au chercheur de faire une découverte inattendue d'importance ou d'un intérêt supérieur à l'objet de sa recherche initiale, et de l'aptitude de ce même chercheur à saisir et exploiter cette chance. Le terme que nous venons d'évoquer et de définir pourrait, dans une certaine mesure, s'appliquer au secteur de l'intelligence artificielle depuis l'irruption de l'apprentissage profond. L'une des caractéristiques essentielles des avancées réalisées dans le secteur de l'IA depuis 2012 réside précisément dans les liens particuliers qui unissent la recherche fondamentale et la recherche appliquée. Actuellement, les progrès successifs auxquels les équipes de chercheurs en IA sont parvenus découlent pour l'essentiel de la recherche appliquée et ont été rendus possibles par des percées techniques. Sur le plan théorique, il existe toujours une difficulté, voire une impossibilité dès lors qu'il s'agit de poser la ou les bonnes questions. En d'autres termes, les avancées d'ordre technique dépassent de loin le rythme de développement de la recherche fondamentale et de la théorisation. On sait comment faire mais on peine à élaborer une théorie qui puisse jeter les bases de nouvelles avancées issues des précédentes. Cette difficulté de théorisation, cette incapacité (souvent temporaire) à formuler une règle générale et abstraite de fonctionnement, ce tâtonnement permanent peut représenter un risque pour le champ de l'IA. L'industrie, qui se situe à la source des principaux financements des réalisations contemporaines de l'IA, pourrait décider de couper les vivres si la recherche n'était pas en mesure de délivrer à moyen terme les « recettes » de futures réalisations. D'une certaine façon, comme le laissait entendre Yann Le Cunn, on n'est jamais à l'abri d'un nouvel hiver de l'IA.

Il convient de comprendre à quel point, durant ces dernières années, la rapidité dans la succession des avancées accomplies dans le domaine de l'IA constitue un faux-semblant derrière lequel se dissimulent des dynamiques de recherche dont les origines remontent à plus de vingt ans. Les récentes percées en matière d'apprentissage profond, souvent attribuées à des groupes privés (Google, Facebook ou

[25] Yoshua BENGIO, « La révolution de l'apprentissage profond », *Pour la Science Hors-Série*, Big Data, numéro 98, février 2018.

Amazon, résultent pour la plupart de progrès technologiques rendus possibles par des travaux de recherche fondamentale conduits il y a près de deux décennies grâce à des financements publics. Cette réalité est parfaitement comprise par les grands groupes qui, au travers de leurs organismes de recherche dédiés à l'IA, maintiennent le lien avec la recherche fondamentale conduite dans des centres de recherche et laboratoires publics. Les résultats des recherches menées par l'ensemble des acteurs de l'IA sont présentés annuellement lors de deux événements majeurs : la première est l'International Conference on Machine Learning (ICML), la seconde est la conférence NIPS (Neural Information Processing Systems). Régulièrement, des scientifiques des principaux groupes privés du numérique tentent d'y faire connaître leurs avancées. Beaucoup y parviennent, mais la compétition des idées est de plus en plus difficile, tant le rythme des avancées proposées est vertigineux. Le secteur de l'IA est, en effet, caractérisé ces dernières années par une hyper-compétitivité.

Il n'en demeure pas moins que, dans l'état actuel, la recherche fondamentale qui devrait permettre aux machines de demain de raisonner (et dépasser le « simple » calcul algorithmique) semble se situer devant un mur insurmontable. Les chercheurs « débroussaillent en espérant trouver les traces d'un sentier qui mènera peut-être vers le Graal. Et personne n'est d'accord sur la direction où chercher. Demis Hassabis, le boss de Google DeepMind, par exemple, ne semble jurer, pour résoudre l'intelligence, que sur l'apprentissage par renforcement. Cette méthode, plus souple que l'apprentissage supervisé, est au cœur d'AlphaGo, le programme qui a défait plusieurs champions de go lors de rencontres très médiatisées. Mais c'est aussi une méthode coûteuse en ressources et qui est pour l'instant limitée à un certain type d'apprentissage. »[26] Ce constat nous oblige à relativiser l'impression d'inéluctabilité pouvant se dégager des avancées récentes engrangées dans le domaine de l'IA et, plus spécialement, de l'apprentissage machine. La recherche appliquée dans ce secteur n'est pas à l'abri de nouveaux ralentissements qui pourraient influer sur la survenance de ruptures technico-militaires à long terme. Au final, si un tel scénario devait se confirmer, il serait à craindre que de nombreuses nations parviennent à se doter des technologies les plus performantes dans le domaine de l'IA et que cette dernière exprime son pouvoir égalisateur en termes de puissance. Le seul obstacle à ce scénario résiderait dans la disponibilité de la donnée.

La *donnée* se situe donc au cœur d'une concurrence mondiale dont peu de monde soupçonne l'existence. C'est pourtant la disponibilité de la *donnée*, quelle que soit son origine ou sa « matière constitutive », qui est pour l'heure l'enjeu principal de la compétitivité des systèmes d'IA à travers le monde. Sans *données*, une IA ne peut apprendre. Et pour apprendre, la quantité de *données* nécessaire ne présente aucune limite. En d'autres termes, quel que soit le niveau et la quantité de *données* qu'il peut obtenir et contrôler, un acteur dominant dans le secteur de l'IA sera toujours en quête d'un montant toujours plus important de *données* pour l'apprentissage de ses systèmes d'IA ; il ne sera jamais rassasié. L'obtention d'une quantité importante de *données* ne garantit toutefois aucunement l'accès à un ensemble qualitatif de données. La conséquence de ce constat est que, selon une logistique probabiliste, plus un acteur disposera de *données* pour alimenter son IA apprenante, plus les chances de disposer d'une *donnée* de qualité seront grandes. Dans un contexte militaire, la profusion de données issues des différentes dimensions de l'espace opérationnel (air, terre, mer, espace et cyberespace) a certes pu représenter un atout fondamental dans la recherche de la supériorité informationnelle sur l'adversaire, mais il est très vite apparu qu'un flux trop important de données recueillies par les divers senseurs déployés pouvait constituer un frein à la prise de décision. En 2006, Thibodeaux & al. ont conceptualisé l'exploitation de l'information dans un contexte militaire à l'aide de la formule suivante : IS = IM+ISR+IO. Cette formule exprime parfaitement l'ensemble des dimensions d'une exploitation informationnelle militaire : la supériorité informationnelle (IS) constitue

[26] Enki BILAL et al., *op. cit.*, p. 18.

la somme de la gestion de cette information (IM), des capacités de renseignement (ISR) et des opérations conduites sur base d'information (IO)[27]. En d'autres termes, la supériorité informationnelle ne découle pas uniquement de la quantité de données issues de l'environnement et recueillies par les capteurs embarqués. Elle suppose, en outre, un travail effectué sur l'information, sa distribution et son traitement mais aussi une classification opportune découlant de la dimension dont les données sont issues. La formule de Thibodeaux & al. se rapporte donc à un processus qui, en raison de la quantité de données recueillies et de la vérification nécessaire de la qualité des données, ne peut être assuré par les seules instances humaines.

La principale difficulté avec le recueil et le traitement de la *donnée* est que ces opérations impliquent des intervenants humains et non humains. Ce ne sont pas seulement des machines et des programmes qui traitent les données et en extraient une décision. La conception même des algorithmes et des méthodes de recueil et de traitement de l'information est le fruit d'un processus humain. L'homme décide du type de données à recueillir, de la manière de les recueillir ainsi que des critères qui détermineront les données valables et exploitables. Dans le cadre de chacune de ces opérations, de nombreux biais – volontaires ou non intentionnels – peuvent interférer. Ces biais sont de nature diverse :

- le biais d'*ancrage* survient lorsque des opérateurs procèdent à une évaluation découlant des toutes premières données recueillies en accordant une importance moindre à la masse totale des données obtenues ;
- le biais de *confirmation* apparaît lorsque des opérateurs interprètent des données en fonction des hypothèses sur lesquelles ces données semblent s'aligner. Il s'agit de l'un des principaux biais cognitifs susceptibles d'affecter la prise de décision. Les mécanismes de dissuasion – conventionnels ou nucléaires – relèvent presque exclusivement de tels biais cognitifs ;
- le biais de l'*interprétation à portée de main* intervient dans une chaîne de décisions lorsque l'opérateur interprète des données sur la base de connaissances personnelles auxquelles il peut se référer de manière immédiate (sans prise en considération de connaissances émanant d'autres opérateurs ou de nature plus globale) ;
- le biais du *train en marche* se rapporte à l'attitude d'un opérateur qui interprétera des données en s'alignant sur les interprétations d'autres opérateurs.

De tels biais sont des phénomènes naturels et récurrents sur le plan cognitif. Ils peuvent intervenir à de multiples stades du processus décisionnel, de l'étape de recueil d'information à celle de la production de la décision même. Il est important pour toute organisation militaire de réduire au maximum le nombre de biais ainsi que leur impact sur le processus décisionnel. Il est souvent affirmé, de manière quelque peu réductrice, que les systèmes d'IA pourraient constituer non pas des succédanés mais des adjoints essentiels à la bonne qualité de la *donnée*, de son traitement et de son intégrité tout au long du processus. Cette affirmation omet le fait que les algorithmes sont toujours conçus par l'homme et qu'ils peuvent intégrer des filtres déformant dans le traitement de la *donnée*. Les données, les recommandations qui en sont issues, les analyses prédictives qui s'appuient sur elles ainsi que les décisions qui résultent des algorithmes de traitement peuvent certes renforcer les capacités d'évaluation d'une organisation militaire tant dans un contexte de crise que dans un contexte de veille situationnelle. Il n'en demeure pas moins que cette « assistance algorithmique » n'est pas à

[27] Maxwell THIBODEAUX, Richard KAPLAN, Anthony SMITH, Joe K. CLEMA, *A Framework for Understanding the IO: C4ISR Relationship*, Colloque « Command and Control Research Program », papers/062, 2006, pp. 1-30.

l'abri de travers cognitifs humains puisque c'est l'homme qui demeure l'architecte des algorithmes[28]. Un tel risque de « défaillance » est, d'une certaine façon, la garantie que l'homme demeurera toujours le responsable ultime d'une dysfonction d'un algorithme. Sur le plan militaire, l'identification claire d'une responsabilité humaine s'avère indispensable au regard du droit et, plus spécifiquement, du droit international humanitaire (DIH).

c) Vulnérabilité de la donnée

Comme nous le mentionnions précédemment, l'accès à la *donnée* constitue désormais le cœur d'une compétition globale entre les États disposant de systèmes d'IA pour leurs besoins en matière de défense. Les volumes d'échanges de données entre particuliers à l'ère numérique atteint des niveaux jamais égalés. Cette dépendance aux données issues de sources toujours plus nombreuses et hétéroclites (la 5G et les objets connectés accroîtront encore de façon géométrique cette tendance) contribue à l'ouverture de brèches et de failles en permettant à des informations tronquées ou falsifiées d'intégrer les bases d'apprentissage des systèmes d'IA. On parle alors de l'empoisonnement de la *donnée* : la manipulation (soit par omission, soit par subtilisation, soit encore par remplacement à l'aide de données corrompues) ou la diffusion involontaire de biais.

Parce que les données avec lesquelles les systèmes d'IA sont aujourd'hui éduqués proviennent des activités des utilisateurs humains, de nombreux biais intègrent les inférences produites par les systèmes d'IA. La fiabilité de la donnée est actuellement au cœur des travaux de nombreux laboratoires privés et publics. Il convient de comprendre que, de nos jours, l'essentiel des données existantes se trouvent au sein d'ordinateurs et de systèmes de stockage et non dans des cerveaux humains. Globalement, l'accroissement de la quantité de données dans le monde s'est opéré selon un rythme exponentiel, largement supérieur, au demeurant, à la loi de Moore. À l'inverse, la capacité humaine à traiter ces données est restée constante. Le corollaire de ce constat est simple : l'homme n'est depuis longtemps plus en mesure d'analyser et de traiter l'ensemble des données qu'il produit.

Déléguer aux machines le traitement de ces données n'est pas chose aisée. Il faut, en effet, que des humains programment des machines pour cette tâche avant même d'envisager d'éduquer ces dernières à l'analyse des données. Or, dans ce travail, il est inévitable que des données que nous qualifions de « biaisées[29] » s'immiscent dans ce processus. Les ingénieurs en charge du développement et de l'éducation des systèmes d'IA ont pris conscience des déformations susceptibles d'affecter les données. Certes, de telles « déformations » ne sont pas uniquement involontaires, elles peuvent également être le produit de techniques spécifiquement destinées à les pervertir. L'avènement de systèmes d'IA en mesure de détecter de tels biais serait proche. Pour les experts, la solution à de tels biais est de tendre vers plus d'IA. Dans la vision du monde qu'ont les développeurs de Google, Facebook, Amazon, etc., l'homme demeurera inévitablement en proie à ses préjugés. La seule manière de corriger les jugements humains et de les épurer des biais cognitifs réside dans l'adjonction de systèmes d'IA ayant été éduqués à leur suppression.

C'est aux alentours de 2011 que des progrès radicaux dans chacun des trois domaines applicatifs cités se produisent. Ainsi, les réseaux de neurones convolutifs sont devenus opérationnels et garantissent des résultats bien au-delà de ce qu'il était alors possible d'obtenir avec les précédentes générations

[28] Lydia KOSTOPOULOS, *The Role of Data in Algorithmic Decision-Making: A Primer*, United Nations Institute for Disarmament Research (UNIDIR), 2019, http://www.unidir.org.

[29] Une donnée en soi ne comporte aucun biais. Elle est simplement la résultante d'un comportement humain par rapport au réel. La donnée intègre donc les jugements de valeur, les perceptions de celui qui la produit. La notion de « biais » devrait plutôt être comprise comme un défaut de « rationalisation » ou de « nettoyage ». Mais là encore, toute la difficulté réside dans la question de savoir comment une telle rationalisation pourrait être opérée indépendamment de tout système de valeur.

de méthodes d'apprentissage profond. C'est l'augmentation de la puissance de calcul des systèmes, rendue possible par l'arrivée des processeurs graphiques, qui a permis aux réseaux de neurones convolutifs, friands en capacités et en énergie, d'être opérants. En d'autres termes, tous les obstacles qui avaient pu exister jusque-là à l'extension des performances de l'IA semblent avoir été levés. L'IA a désormais cessé d'être cantonnée à une branche discrète de l'informatique pour s'ériger en nouvel écosystème.

La question qui peut être posée aujourd'hui est de savoir si nous évoluons indiscutablement vers l'émergence d'une IA forte. Aux dires de quelques experts et technologues, l'avènement d'une IA générale ne serait qu'une question de temps, essentiellement liée à la dynamique d'évolution des récentes ruptures technologiques intervenues dans le domaine des processeurs graphiques et des réseaux de neurones. C'est notamment la position tenue par Ray Kurzweil. Pourtant, si des progrès réels et indéniables semblent être intervenus en la matière, il convient de rappeler qu'aucune IA n'est aujourd'hui en mesure d'expliquer les causes de ce qu'elle observe et analyse. Elle est – et demeurera sans doute pour longtemps – tributaire de la fiabilité des données que l'homme lui soumet pour évoluer. Cette dépendance constitue un biais critique pour n'importe quelle IA.

L'attention médiatique que suscitent parfois les prouesses de l'IA doit être relativisée. Battre un humain lors de parties d'échecs ou de jeu de go ne permet en rien d'anticiper les capacités futures d'une intelligence artificielle dans d'autres domaines parfois variés. Les systèmes experts, certes performants et objets de progressions géométriques, demeurent malgré tout des systèmes experts, consacrés à l'accomplissement supervisé de tâches spécifiques et n'attestent en aucune façon de l'émergence à moyen ou long terme d'une intelligence artificielle forte ou générale.

3. L'IA comme enjeu d'un discours sociopolitique

Au-delà des ruptures technologiques dont sa progression est le produit, l'IA s'inscrit également au cœur d'un discours social et politique. Selon les plus ardents défenseurs de l'IA, celle-ci incarnerait la solution par excellence à la complexité du monde dans tous ses aspects. En d'autres termes, il serait difficile d'imaginer à l'avenir une administration des affaires publiques et privées se passant d'IA. La réalité est plus nuancée. Luc Julia, cocréateur du logiciel Siri (sur les appareils de la société Apple) affirme que l'intelligence artificielle n'existe pas. Andrew Moore, vice-président de Google, affirme quant à lui que « l'IA est actuellement très, très stupide », Laurent Alexandre prétendait qu'en 2017, l'IA était foncièrement *inintelligente*. Sans verser dans de telles extrémités déclaratoires, il convient en effet de comprendre que nombre d'idées relatives aux perspectives de l'IA relèvent pour l'essentiel du mythe (nous l'avons déjà dit) véhiculé par nombre de visions fantasmagoriques à propos des promesses et déboires de la technologie.

Ainsi en va-t-il de la « prétendue » autonomie conférée par l'IA. S'il est exact que l'intelligence artificielle consiste dans le développement d'algorithmes complexes autorisant à des dispositifs d'évoluer dans des configurations où seul l'homme est pour l'instant en mesure de réagir avec efficience, cela ne signifie pas pour autant qu'il s'agisse là d'un système réellement intelligent ou autonome. Comme le rappelle Patrick Bezombes, « *l'autonomie, tout comme l'intelligence, est une des caractéristiques de l'homme qui a acquis un certain savoir-faire et des connaissances. Étymologiquement, l'autonomie consiste à être gouverné selon ses propres règles : l'homme est donc bien autonome, puisqu'il lui est toujours possible de sortir, à ses risques et périls, du cadre qui lui est fixé. Dans de nombreux cas, le non-respect de la règle et la capacité à changer les règles établies dans des environnements nouveaux permettent à l'homme, par un processus quasi darwinien, d'évoluer*. »[30]

[30] Patrick BEZOMBES, *op. cit.*, p. 13.

La faculté de désobéissance est donc, en quelque sorte, le propre de l'homme et découle de son autonomie.

On imagine mal des systèmes informatiques ou autres dispositifs dotés d'une réelle autonomie, c'est-à-dire capables de s'affranchir des règles qui leur ont été fixées pour évoluer dans les environnements pour lesquels ils ont été créés. Un système autonome – a fortiori un système d'armes létal autonome – capable de renoncer à l'application des règles qui lui ont été assignées ne pourrait être toléré par les hommes. Et c'est la raison principale pour laquelle les organisations militaires – dans leur dimension opérationnelle – ne peuvent tolérer l'intégration de systèmes susceptibles de se démarquer des ordres humains. D'ailleurs, les classifications et taxinomies d'un organe comme la très sérieuse National Highway Transportation Safety Authority aux États-Unis, lorsqu'elle envisage la place des véhicules dits « autonomes », n'évoquent pas ces derniers sous cette appellation mais désigne plutôt de tels systèmes sous des termes tels que : *automated vehicles* ou *automated driving systems*. Le débat portant sur ce qui est abusivement labellisé sous l'étiquette de systèmes d'armes létaux autonomes gagnerait immanquablement en sérénité en revoyant une appellation volontairement provocatrice et en rupture avec le réel.

Des réserves similaires doivent être faites au sujet de la capacité d'auto-apprentissage des systèmes d'IA. Certes, l'apprentissage dont il est question ici (notamment depuis l'introduction des processeurs graphiques et le développement des réseaux de neurones) se distingue de la simple « mise en mémoire ». La différence entre ces deux mécanismes réside dans le fait qu'un « apprentissage » implique une modification des règles de comportement de l'IA tandis qu'une mise en mémoire ne suppose qu'un stockage d'information. Il n'en demeure pas moins que, même dans le cas d'un apprentissage, l'IA « comprend » ce qu'elle apprend.

Une IA combine donc « automatisme » et « apprentissage ». Cela signifie donc qu'elle dispose de la capacité de modifier son comportement (et non la règle sur laquelle se base cette possibilité de modification) sur la base de ce qu'elle « apprend ». En réalité, l'algorithme qui se situe au cœur de l'IA applique au problème qu'elle a pour mission de résoudre une opération d'optimisation perçue à tort comme un « apprentissage ». Une IA adopte donc plus une règle qu'elle ne l'adapte.

Le discours de l'IA doit également s'entendre en termes de relations de pouvoir entre les acteurs. L'intelligence artificielle, et surtout les données de masse sur lesquelles elle s'appuie, en est venue à constituer un véritable enjeu de puissance. Pour Laurent Alexandre, il existerait deux catégories d'acteurs en matière de données. La première regroupe les États, et en leur sein les entreprises, qui parviennent par diverses stratégies commerciales à capter et amasser de la donnée. Plus prosaïquement, deux grands acteurs internationaux sont aujourd'hui les principaux collecteurs de données : les États-Unis et la Chine. L'Europe, à l'inverse, est parfois qualifiée de « tiers-monde » de la donnée. De la même façon que les populations du tiers-monde s'extirpaient des conditions tragiques de leurs lieux d'existence en se déplaçant, de manière volontaire ou forcée, vers les pays riches, ainsi les pourvoyeurs de matières premières du XXI^{e} siècle (les données) exportent majoritairement leurs diverses informations personnelles (images, fichiers, vidéos, données biométriques, informations médicales, physiologiques et métaboliques, déplacements, etc.) vers des entreprises étrangères qui les transforment et en retirent la plus grande part du bénéfice. La contrôle des données est *la* clé de voûte de la domination que pourrait conférer demain l'IA.

Aussi, pour s'assurer du transfert toujours plus abondant de données vers les détenteurs de technologies numériques (qui par la même occasion sont également les dépositaires de ces données stockées en masse), un discours de séduction quant aux bienfaits futurs de l'IA est indispensable.

Celui-ci comporte des facettes multiples allant du simple argument de facilitation du quotidien[31] aux envolées post- ou trans-humanistes à propos du devenir de l'individu dans un monde ultra-numérisé où l'intelligence humaine sera dopée à renforts de nano-implants et autres prothèses supposées extraire l'homme de sa condition de simple mortel. Afin de soutenir ces prophéties, de grands cabinets de conseil en stratégie d'entreprise multiplient les rapports d'évaluation d'impact de l'intelligence artificielle sur le marché. L'IA est alors présentée comme créatrice et multiplicatrice de valeurs. Face aux craintes liées aux destructions d'emplois humains que pourrait entraîner avec elle l'intelligence artificielle, le ton des organes de consultance se veut expressément rassurant et optimiste. Le concept de « destruction créatrice » de Schumpeter est invoqué pour certifier que les emplois de demain, non encore existants et par définition non encore fournis, remplaceront les métiers qui d'ici demain auront disparus. De tels discours véhiculent à l'envi une approche déterministe de l'IA. Ils suggèrent une évolution en profondeur de la nature mais aussi de la géographie de la puissance[32].

4. IA faible, IA forte et IA générale

Les concepts d'intelligence artificielle faible et d'intelligence artificielle forte apparaissent de façon récurrente dans la littérature spécialisée. On observe également depuis quelques années l'émergence d'une notion nouvelle : l'intelligence artificielle générale (General Artificial Intelligence, GAI). Une vision quelque peu simplificatrice de la distinction entre lesdites notions revient généralement à dire que l'on désigne par IA faible une intelligence artificielle conçue pour la réalisation de tâches spécifiques. Tous les systèmes d'IA existants sont d'ailleurs considérés comme des systèmes d'IA faibles. Face à elle, existerait donc l'IA forte (parfois qualifiée de générale) que l'on présente habituellement comme une « machine intelligente » capable de résoudre tout type de problème de quelque nature que ce soit. Par extension, il est parfois affirmé que l'émergence d'une IA forte – ou générale – correspondra à l'apparition d'une intelligence artificielle « réelle », dotée d'une conscience propre.

Nous le devinons : une telle présentation ne rend compte en rien de la profondeur réelle du débat.

Les concepts d'IA faible et forte doivent leur existence à un philosophe, John Searle, qui, dans les années 1980, s'attacha à examiner la structure de l'IA en tant que discipline. Il est souvent reproché à John Searle d'avoir voulu porter le discrédit sur la discipline de l'IA dans son ensemble. Il n'en est pourtant rien. Au contraire, les réflexions du philosophe ont permis d'opérer une distinction plus claire entre les activités scientifiques qui relèvent véritablement de l'IA et les visions prophétiques qui entourent la notion d'IA forte. Expliquons. Pour rappel, l'expression « intelligence artificielle » fut introduite en 1955 par le mathématicien John McCarthy qui proposa avec trois autres scientifiques (Marvin Minsky, Nathan Rochester et Claude Shannon) un projet d'école d'été dont le sujet était de concevoir de nouvelles méthodes visant à approcher les facultés cognitives humaines à l'aide de machines. L'objectif poursuivi par les quatre fondateurs de l'intelligence artificielle était d'explorer et de comprendre l'intelligence (humaine et animale) en reproduisant sur des ordinateurs ses multiples manifestations : raisonnement, mémoire, calcul, perception, etc. L'intelligence artificielle était donc, à son origine, une méthodologie, une étude expérimentale s'appuyant sur les nouvelles technologies de l'information de l'époque. À son origine, la discipline de l'IA n'avait aucune ambition démiurgique ; elle entendait, par l'expérimentation, parvenir à une meilleure connaissance des fonctions diverses de

[31] On pensera notamment au discours consistant à banaliser l'omniprésence de l'IA au travers de technologies largement utilisées, telles que la messagerie instantanée, les e-mails, les outils de géolocalisation ou les logiciels de traduction.

[32] Julien NOCETTI, *Intelligence artificielle et politique internationale*, Paris, Institut français des relations internationales, Études de l'IFRI, novembre 2019, p. 14.

l'intelligence humaine. Il ne s'agissait nullement de produire un double de l'homme ou d'envisager le déclassement de l'homme par l'IA.

Les nombreux succès de l'IA « faible »[33] en tant que discipline ont conduit certains scientifiques et philosophes qui ne figuraient pas parmi les pionniers à envisager l'IA sous un angle différent : une sorte de poursuite du processus de rationalisation entamé avec les Lumières. Très rapidement, un parallèle allait être établi par les partisans du cognitivisme entre l'ordinateur et le fonctionnement du cerveau humain. L'idée désormais développée était d'aboutir à une intelligence artificielle forte en s'appuyant sur l'allégation selon laquelle le cerveau humain pourrait être « réduit » à une machine juste plus complexe. Une telle approche cognitiviste de l'IA allait conduire à une scission nette entre, d'un côté, les pionniers de l'IA et, de l'autre, les partisans d'une posture philosophique de l'IA. On ajoutera que cette même période (le milieu des années 1950) assiste à l'émergence de la *cybernétique* sous l'impulsion de Robert Wiener. Tant l'IA que la cybernétique sont les purs produits de la Seconde Guerre mondiale et s'inscrivent dans un contexte de guerre froide opposant deux superpuissances aux capacités de destruction sans précédent. Le projet de développement d'une *machine intelligente* se fit jour parmi les membres de l'école cybernétique. Le projet des cybernéticiens visait le découplage du corps et de l'intelligence. Autrement dit, défaire l'homme des limites de son corps afin d'accéder à un état supérieur, corriger les imperfections de la nature. Dans le modèle proposé par l'école cybernétique, la correction qu'il convient d'opérer s'insère dans un mode de gouvernance spécifique de contrôle technocratique (le terme cybernétique provient du même mot grec qui a donné, par le latin, « gouvernail », « gouvernement », « gouvernance »). Ce type de contrôle est alors très convoité dans une période, la guerre froide, où l'ensemble des forces sociales et économiques, qu'elles se situent dans le camp libéral ou communiste, sont mobilisées autour d'une opposition quasi civilisationnelles entre deux formes de société. Les années 1950 sont également la décennie du maccarthysme et de la paranoïa collective qui frappe, aux États-Unis, les sphères dirigeantes quant à l'infiltration de l'appareil d'État par les communistes. Alors même que l'idée d'un contrôle technocratique de la société semble s'inscrire dans le phénomène de chasse aux sorcières qui se développe aux États-Unis, le projet de conception d'une *machine intelligente* vise aussi à mettre un terme aux dérives de l'esprit humain que furent les atrocités de la Seconde Guerre mondiale et qu'incarnèrent la véritable lutte à mort entre les nations européennes mais surtout l'Holocauste. Une machine à gouverner : tel est en substance le projet de certains tenants de la cybernétique. Cette ambition se rapproche à bien des égards du projet de conception d'une intelligence artificielle forte, capable d'effectuer parmi les nombreuses et diverses tâches qui lui seraient assignées celle d'une gouvernance nouvelle, technocratique, axée sur la régulation de la machine[34].

La notion d'intelligence artificielle générale, quant à elle, a émergé au début du XXI^{e} siècle et ne doit en aucune façon être associée aux travaux de l'IA entamés à la fin des années 1950. L'idée qui se situe derrière la notion d'IA générale (IAG) est de refonder l'intelligence artificielle sur des bases mathématiques solides, dont le degré de certitude des résultats équivaudrait à celui du domaine de la physique. Pour les tenants de l'IAG, l'ensemble des théorèmes mathématiques à la base de la science générale de l'intelligence artificielle étant démontrés, il en découle que la réalisation d'une véritable intelligence artificielle générale ne dépend que de la capacité de stockage et de calcul des machines. Contrairement aux partisans de l'intelligence artificielle forte dont les postures relèvent des champs philosophique et discursif (et permettent en cela d'être clairement distingués des travaux pionniers de

[33] Il importe de rappeler les aboutissements auxquels ont donné lieu l'étude de l'IA « faible ». Parmi ceux-ci, on citera notamment – et de manière non-exhaustive – l'HyperText Markup Language (HTML), la biométrie, la reconnaissance faciale, les moteurs de recherche, le traitement de masse de données, l'apprentissage machine, etc.

[34] À ce sujet, voir Céline LAFONTAINE, *L'empire cybernétique : des machines à penser à la pensée machine*, Paris, Seuil, 2004, pp. 49 et ss.

l'IA), l'intelligence artificielle générale prétend s'appuyer sur des travaux mathématiques contestables (et souvent contestés) ainsi que sur des réalisations informatiques fort discutables. Les extrapolations futuristes découlant de la loi de Moore figurent parmi celles-ci. L'IAG compte aujourd'hui parmi ses partisans de nombreux *singularitariens*, c'est-à-dire des adeptes du principe de *singularité*.

La loi de Moore

Il existe, en réalité, deux lois de Moore exprimées par Gordon E. Moore lui-même dans la revue *Electronics Magazine*. Ces lois « empiriques » expriment plus précisément des conjectures. Plusieurs autres lois sont venues s'adjoindre aux deux lois originelles et ont été erronément assimilées à celles-ci. La première loi de Moore procède du constat selon lequel la complexité des semi-conducteurs proposés en entrée de gamme doublait tous les ans à coût constant depuis 1959, date de leur invention et mise sur le marché. Gordon Moore postulait donc la poursuite de cette croissance. La seconde loi de Moore, exprimée en 1975, constitue au vrai une correction de la première au vu des réalisations obtenues depuis 10 ans. Cette seconde loi, qui n'est autre, ici encore, qu'une extrapolation empirique, soutient que le nombre de transistors des microprocesseurs (on ne parle plus de circuits intégrés) sur une puce de silicium double tous les deux ans. Cette prédiction se révélera par la suite d'une extraordinaire exactitude dans les faits. Une troisième loi, erronément attribuée à Gordon E. Moore, postule que la puissance (on parle aussi, sans différenciation, de capacité ou de vitesse de calcul) des microprocesseurs double tous les dix-huit mois. Cette prédiction n'est, en réalité, qu'une généralisation imparfaite – sinon fausse – des première et seconde lois de Moore. Elle est, cependant, l'expression de la loi de Moore la plus véhiculée.

5. Quelles perspectives pratiques pour une IA militaire ?

La question à laquelle nous tenterons de répondre ici consiste à établir une liste, certes non exhaustive (et non dépendante des partisans), de l'apport de l'IA dans le domaine militaire. D'une manière générale, l'IA présente plusieurs atouts :

- la résolution de problèmes complexes et non déterministes ;
- la détection, la caractérisation et le traitement automatiques d'informations dans des quantités et dans des délais de traitement avec lesquels l'homme ne peut rivaliser ;
- l'établissement de corrélations entre des domaines non connectés à l'aide de données issues de milieux hétérogènes.

Toutefois, ces divers avantages peuvent s'avérer en opposition avec les contraintes mêmes de l'environnement militaire. En effet, comme cela a été mentionné précédemment, les progrès récents réalisés dans le domaine de l'intelligence artificielle découlent principalement de l'accumulation sans précédent de données résultant de l'emploi fait des technologies numériques par les utilisateurs. Ce sont ces gigantesques bases de données les plus diverses qui ont alimenté les processus d'apprentissage (*deep learning*) des systèmes d'IA. Or, afin de pouvoir effectuer des missions de nature militaire, les systèmes d'IA conçus pour ces missions doivent pouvoir compter sur des données issues de l'environnement opérationnel militaire et de leurs « utilisateurs ». Cette exigence se heurte toutefois à plusieurs obstacles, tels que :

- les différents types de classification des données établis afin de restreindre l'accès à des informations jugées sensibles ;
- la confidentialité même de certaines données ;
- les différentes sphères d'appartenance des données recueillies (les données sont souvent récoltées par services, armes, etc.) ;

- l'insuffisance des échanges de données et d'informations sur les théâtres d'opération entre différents pays d'une même coalition ;
- la consommation en énergie des systèmes d'IA embarqués[35].

On le voit : à l'instar de toute rupture technologique, l'IA impose de trouver un certain équilibre entre les possibilités nouvelles offertes par la technique et les contraintes résultant de l'environnement « spécifique » dans lequel elle est supposée déployer ses apports. Encore faut-il déterminer les domaines précis dans lesquels l'IA pourra faire montre d'un avantage comparatif.

6. L'IA : une technologie déjà fort répandue dans nombre de secteurs

La robotique, à laquelle elle est souvent associée, n'est pas le seul domaine exploitant les avancées de l'intelligence artificielle. Celle-ci se retrouve dans des objets et des systèmes très courants, qui concernent aussi bien les activités marchandes que le grand public.

Bien des secteurs dans le milieu civil et commercial recourent aujourd'hui à l'IA. On peut citer notamment le domaine des banques, des assurances et les institutions financières. Ainsi, par exemple, la plupart des prêts accordés par les établissements bancaires sont accordés sur la base d'une analyse de profil client assistée par un système d'IA. Ce sont encore des logiciels experts qui prodiguent des conseils en matière de placements financiers et boursiers. Enfin, et s'il ne fallait citer qu'un seul exemple où l'homme a complètement cédé la place à la machine, c'est sans nul doute le domaine des transactions à haute fréquence (ou trading haute fréquence – THF).

Le secteur médical fait également figure de domaine pionnier pour un certain nombre d'applications en intelligence artificielle. Ces applications pourraient se révéler particulièrement prometteuses et déboucher sur des perspectives de développement intéressantes dans le secteur militaire, notamment en matière de chirurgie réparatrice pour des cas de mutilations de guerre. Les systèmes d'IA employés dans le secteur médical disposent, en effet, d'une capacité de croisement de résultats d'examens, de diagnostics et de littérature scientifique absolument sans comparaison avec l'expertise d'un être humain formé à la médecine, même la plus spécialisée. Par ailleurs, en matière de prescriptions médicamenteuses et de risques d'interaction entre molécules, les systèmes d'IA s'avèrent plus performants pour la lutte contre les effets secondaires liés à la combinaison de plusieurs traitements. Chaque patient peut ainsi espérer bénéficier d'un traitement ultra-personnalisé quelles que soient les pathologies à traiter.

Enfin, parmi les secteurs les plus consommateurs d'IA, on signalera le *marketing*. Les technologies de l'IA rendent possible un meilleur affinement des sollicitations commerciales à l'adresse du consommateur. Afin d'éviter que celui-ci ne ressente une pression trop forte – ce qui pourrait l'amener à couper tout contact – des algorithmes définissent, sur la base des historiques d'achat et des habitudes en matière de courriers, le meilleur moment pour l'envoi de publicités. Certaines entreprises ont également conçu des systèmes de messagerie publicitaire ciblée qui s'appuient sur l'enregistrement et le traitement en temps-réel des échanges entre les clients et les services après-vente pour transformer ceux-ci en données structurées et immédiatement exploitables. On ne manquera pas, enfin, d'évoquer le développement d'assistants *shopping* qui prennent en compte les habitudes de dépense du client.

[35] On peut, toutefois, imaginer que le processus d'auto-apprentissage des systèmes d'IA s'effectuerait à partir de dispositifs non-embarqués.

7. Conclusion partielle

Expression on ne peut plus ambivalente, l'intelligence artificielle fut déclinée en de nombreuses variantes depuis son émergence au lendemain de la Seconde Guerre mondiale. L'IA apparaît comme le résultat de la recherche militaire qui appuya l'entreprise scientifique du projet Manhattan et la conception de la première bombe atomique. Qualifiant à ses débuts un champ d'étude consacré à la compréhension des fonctions cognitives humaines, l'intelligence artificielle désigna par la suite certaines matérialités supposées reproduire le raisonnement humain au travers d'artefacts. L'IA a vu sa filiation avec la cybernétique s'estomper au fil du temps pour évoluer vers la conception de systèmes. Les progrès de l'IA n'ont pas suivi une voie linéaire, mais ont procédé par bonds et reflux. Les « hivers de l'IA » ont pu conduire à interroger la viabilité des projets portés par la discipline. Toutefois, les évolutions récentes intervenues dans le domaine de l'apprentissage machine profond et automatique ainsi que les travaux menés en matière de réseaux de neurones artificiels laissent entrevoir des perspectives d'applications nouvelles, plus particulièrement à finalité militaire. Il n'en demeure pas moins que le champ des réalisations promises s'avère aussi vaste que celui des questionnements qu'il comporte.

II. Quelles potentialités militaires pour l'IA ?

L'intelligence artificielle présente des atouts multiples pour le stratège et le chef militaire dans le cadre de leurs responsabilités opérationnelles et organisationnelles. Il importe donc pour les organisations de défense d'être en mesure d'extraire de l'IA toutes les potentialités technologiques et de transformer celles-ci en des facteurs susceptibles de se révéler décisifs pour atteindre une supériorité dans l'espace de bataille. L'apport de l'IA se fera notamment sentir par un gain en terme de vélocité, par une marge de manœuvre plus étendue, par une meilleure connaissance de l'espace de confrontation (grâce à une meilleure capacité de reconnaissance et de détection des cibles) et par des actions plus rapides et mieux ciblées.

1. La compréhension et l'anticipation

Des modalités nouvelles de traitement des données, conjuguant l'analyse croisée et massive, sont désormais rendues possibles par l'IA. Les capacités que confèrent l'IA dépassent de loin les aptitudes de traitement humaines. On peut donc raisonnablement s'attendre à ce qu'une compréhension plus complète et plus rapide des situations de crise et des espaces opérationnels (plus complexes et interdépendants) soit désormais possible. L'intelligence artificielle contribuera à une meilleure anticipation des manœuvres adverses et permettra une optimisation des processus opérationnels dans l'ensemble des segments que sont l'orientation, la recherche, l'exploitation et la diffusion du renseignement. La rapidité d'accès à l'information d'un théâtre de crise autorisera un élargissement considérable de la fenêtre de temps accordée à l'examen des options opérationnelles et à l'étude des hypothèses d'action pouvant être prises. Peut-être la contribution essentielle des technologies de l'IA se situera-t-elle dans l'examen des signaux faibles qui, auparavant, passaient sous la « couverture radar » des services de renseignement. Une meilleure récolte et mise en perspective des signaux faibles, analysés en temps-réel et selon des boucles itératives, réduira considérablement la possibilité pour l'adversaire de mise sur l'effet de surprise. De telles aptitudes rendues possibles par l'IA assureront des gains de tempo opérationnel et donc la possibilité pour les forces armées de disposer d'un ascendant sur les manœuvres de l'adversaire.

2. L'apprentissage autonome et en partage

Le secteur dans lequel une IA militaire pourra faire montre d'un avantage comparatif certain est sans nul doute le partage des processus d'apprentissage. Afin de garantir son opérationnalité sur un théâtre d'opération ou dans un processus décisionnel politico-militaire, une IA devra être en mesure d'apprendre de situations les plus diverses se présentant au cours d'une intervention de quelque type que ce soit, afin d'affûter au mieux sa capacité de décision en des circonstances précises.

Il convient, au préalable, de revenir quelques instants sur la notion d'apprentissage dans le secteur de l'IA. Contrairement à une idée fort répandue – souvent véhiculée par une certaine mystification des technologies –, un système d'IA ne procède pas à un auto-apprentissage mais plutôt à un apprentissage automatisé. Ce type d'apprentissage va donc au-delà de la simple « mémorisation » puisque le processus implique une modification des règles de comportement de l'IA. Ici encore, arrêtons-nous quelques instants sur cette notion de « modification des règles de comportement ». Il ne s'agit pas d'une modification des règles qui régissent l'environnement dans lequel évolue l'IA, mais bien de celles qui président à son interaction avec l'environnement dans lequel elle devra évoluer et conduire un certain nombre d'actions[36]. En d'autres termes, un système d'IA « apprenant » alimente

[36] Patrick BEZOMBES, « Intelligence artificielle et robots militaires », *Défense & Sécurité Internationale*, Hors-Série, numéro 65, avril-mai 2019, pp. 12-15.

son algorithme de fonctionnement en vue d'optimiser son « niveau de jeu » en fonction d'un contexte. Un système d'IA apprenant n'est donc, au final, qu'un agent computationnel.

La question qui demeure toutefois est la suivante : un système d'IA apprenant peut-il, au-delà de l'optimisation de son comportement face à son environnement, redéfinir les règles mêmes qui guident ce processus d'adaptation ? Plus encore, un tel système d'IA est-il en mesure de redéfinir de telles règles en faisant fi d'une validation humaine ? Sur un plan strictement technique et matériel, des systèmes d'IA dotés de telles facultés existent déjà. Ce sont de tels modes d'apprentissage qui ont permis le développement des *chatbots* et donné lieu à des comportements adaptatifs déviants[37].

La tentation de permettre de telles formes d'apprentissage à des systèmes d'IA, c'est-à-dire sans supervision ni validation humaine, est particulièrement forte dans certains milieux de la R&D consacrée à cette discipline. Cette perspective s'avère particulièrement profitable à des entreprises qui souhaiteraient s'affranchir des coûts importants liés aux modalités de contrôle et de validation des processus d'apprentissage de leurs systèmes d'IA. Un tel abandon donnerait, par ailleurs, la possibilité pour tout système d'IA en mesure d'évoluer de la sorte de traiter un nombre considérable – quasi infini – de données, et ce en temps réel ou quasi réel.

Enfin, et c'est là que se situe l'un des immenses avantages de l'IA sur la formation des humains, une machine auto-apprenante pourra procéder à un partage immédiat des optimisations de ses algorithmes à d'autres systèmes d'IA dotés de fonctions logiques identiques. Cette possibilité assure un gain de temps considérable en matière de formation et, surtout, des économies d'échelle considérables au niveau des coûts.

3. La personnalisation

Autre intérêt de l'IA dans le domaine militaire, la personnalisation des équipements du combattant et l'ajustement de dispositifs informatiques d'aide assureront une plus grande agilité tant physique qu'intellectuelle dans les prises de décision et modalités d'action. On peut, ainsi, évoquer le développement d'interfaces homme/machine sur la base de profils personnalisés résultant des interactions répétées entre le combattant et son matériel au gré des habitudes constatées et des souhaits spécifiques de l'individu. On mentionnera encore la mise au point d'interfaces cognitives facilitant l'expérience de son utilisateur pour la prise de connaissance des données issues du contexte dans lequel évolue le combattant. Les informations transmises seront ainsi optimisées et continuellement adaptées pour fournir une représentation mentale de l'environnement qui soit de compréhension facile et propre à des décisions au plus proche de la réalité du terrain. L'IA permet également une plus grande adaptation du soldat à son environnement socio-culturel. Lorsqu'il est amené à progresser dans des zones étrangères, des traducteurs instantanés reposant sur l'IA peuvent contribuer à une interprétation en temps réel avec les populations indigènes.

[37] Barthélémy DONT, « Amazon a dû se débarrasser d'une intelligence artificielle sexiste », *Slate*, 10 octobre 2018, *cf.* http://www.slate.fr/story/168413/amazon-abandonne-intelligence-artificielle-sexiste. Dans cet article, l'auteur rapporte que l'intelligence artificielle de la société censée procéder de manière objective (et sans a priori sexiste du fait de sa nature informatique) à la sélection des meilleurs et candidates en vue de pourvoir des postes vacants au sein de l'entreprise avait tendance à écarter davantage les candidatures de femmes et retenir en majorité des candidatures masculines. L'IA, bâtie sur un processus de *deep learning* constitué de données préexistantes issues de précédentes campagnes de recrutement, avait reproduit les biais humains des jurys d'embauche de l'entreprise. Voir aussi Carole CRIADO-PEREZ, *Invisible Women: Exposing Data Bias in A World Designed for Men*, Vintage Publishing, 2019.

4. Le traitement optimisé des données

Lors d'opérations militaires, l'une des problématiques rencontrées par le combattant est la surabondance d'informations relatives au théâtre d'opération dans lequel il est amené à progresser. Cette surcharge de données, dont une part parfois considérable ne s'avère pas essentielle à la mission, peut handicaper la prise de décision aux niveaux tactique, opérationnel et stratégique, voire même paralyser l'action militaire dans des situations où celle-ci peut s'avérer nécessaire. L'extraordinaire multiplication et diversification des senseurs (satellites, drones, senseurs terrestres, maritimes), des sources de données (HUMINT, cyberespace, organes diplomatiques, informations *open source*, etc.), ainsi que l'étendue du spectre de données (radar, infra-rouge, etc.) appelle à adopter des modes de traitement et d'analyse performants, l'homme n'étant plus en mesure d'opérer la fusion des données recueillies. Ce travail d'agrégation et de corrélation recourt désormais à l'IA. Celle-ci permettra la détection de signaux faibles, souvent précurseurs d'évolutions sécuritaires que ne peut repérer un traitement des données par l'homme. L'apport de l'IA, du fait de sa réactivité, résidera aussi dans la détection d'événements dans le cyberespace et de leur impact sur l'évolution d'une situation de crise (*fake news*, rumeurs, propagande, campagnes de désinformation, conjonctures économiques, sociales, culturelles, religieuses, etc.) : un faisceau d'une multitude de données provenant de sources extrêmement diverses (parfois sans rapport, en apparence, avec une situation de crise ou un théâtre d'opération) mais dont les répercussions, du fait de la « collision » de certains événements, pourraient se révéler critiques dans la conduite de l'action stratégique.

5. Planification et conduite

L'apport de l'IA pourrait se révéler décisif en matière de préparation et d'automatisation des vecteurs. Les systèmes d'IA permettront, en effet, une modélisation des scénarios d'opérations fondées sur un nombre considérable de variables (disponibilité des ressources humaines et matérielles, conditions de terrain, déploiement des forces, etc.). Sur cette base, ils assisteront la planification d'opérations en amont, afin de tester les différentes options pour offrir aux décideurs un éventail de possibilités. Dans le cadre opérationnel, de tels dispositifs d'assistance par IA pourront également procéder à une réévaluation permanente des plans de mission afin d'adapter ceux-ci au mieux et en temps réel aux données nouvelles issues de l'espace de bataille[38].

L'IA peut aussi être envisagée pour le déploiement automatisé (et non pas autonome) de certaines catégories de systèmes d'armes. Ainsi en est-il de l'*Iron Dome* israélien, un système mobile de défense antimissile de couche basse destiné à protéger une zone de 70 km contre les missiles de très courte à courte portée et les obus d'artillerie de 155 mm. Après avoir repéré l'objet suspect dès son lancement, l'*Iron Dome* calcule la trajectoire du projectile et détermine s'il convient de l'abattre ou non en fonction de la menace concrète qu'il fait peser sur la sécurité des populations, villes et forces armées. Cette décision est prise de manière autonome et quasi instantanée.

6. Drones et plates-formes autonomes

L'IA présentera un intérêt fondamental dans la gestion des systèmes semi-autonomes et autonomes. L'utilisation de drones en essaim (*swarming drones*) s'appuiera inévitablement sur des technologies d'IA pour assurer la coordination de chaque unité dans l'ensemble déployé. Soit il s'agira d'un pilotage à distance depuis un système d'IA via des réseaux satellitaires relais, soit l'essaim comportera en son cœur un « engin-mère » doté d'un système d'IA embarqué pour le pilotage des diverses unités.

[38] Axel DYÈVRE, Pierre GOETZ, Florence FERRANDO, *Intelligence artificielle : applications et enjeux pour les Armées*, Paris, Compagnie européenne d'intelligence stratégique (CEIS), coll. Les Notes Stratégiques, septembre 2018.

Le déploiement de mini-drones en essaim constituera un atout essentiel pour l'exécution de diverses missions : saturation des défenses antiaériennes ennemies, recueil du renseignement selon un mode collaboratif sur de larges zones (éventuellement des zones contaminées), etc. Aux États-Unis, le Pentagone a procédé au déploiement d'un essaim de 103 mini-drones *Perdrix* lancés depuis trois avions de chasse. De la même façon, la Royal Air Force a testé le principe d'une escadre de *swarming drones* essentiellement destinés à saturer les défenses ennemies[39].

Illustration 1 : Mini-drone *Perdrix* déployé par l'U.S. Air Force

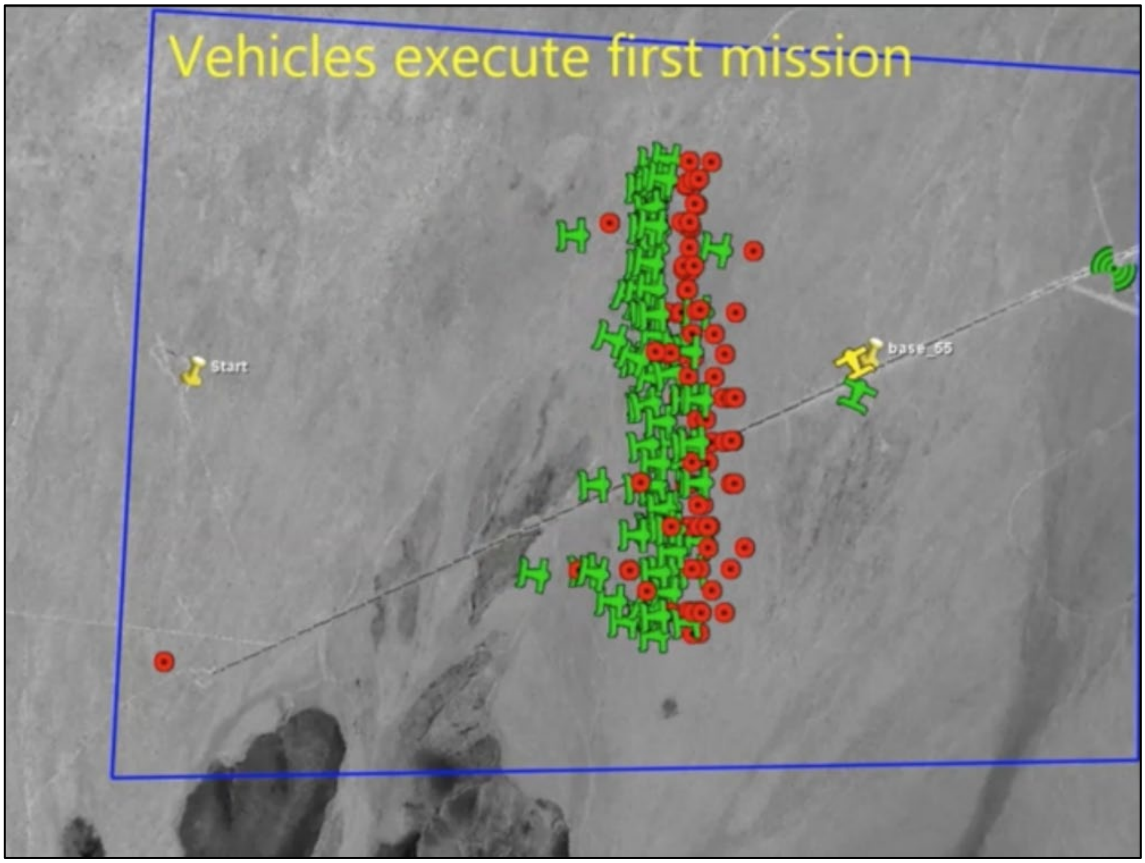


Illustration 2 : essaim de 103 mini-drones déployé par trois avions de combat américains lors d'une démonstration technologique en octobre 2016

[39] Harry LYE, « RAF to launch swarming drone squadron in April », Air Force Technology, 13 janvier 2020, *cf.* https://www.airforce-technology.com/news/raf-swarming-drones.

7. La protection du soldat

La contribution des technologies de l'IA au bénéfice du soldat s'appuiera sur des apports très variés. Ainsi, l'assistance de l'IA pourra se révéler fondamentale dans le traitement massif et en temps-réel des données de santé provenant des capteurs qui monitoreront le soldat durant son déploiement. Une meilleure identification préalable des facteurs de risques liés aux environnements et aux conditions d'emploi des forces sera également permise par l'appui de l'IA.

En matière d'entraînement, les technologies de l'IA se révéleront des outils essentiels dans la conception de meilleurs simulateurs et, de cette façon, prépareront de façon plus optimale les combattants aux conditions de terrain qu'ils rencontreront. Des simulations dans des environnements virtuels immersifs, très proches de la réalité opérationnelle (sans toutefois l'égaler en tous points) garantiront une meilleure anticipation des risques et des options de déploiement.

Ensuite, associées aux avancées dans le domaine de la robotique, les technologies de l'intelligence artificielle permettront la confection d'assistants plus véloces, adaptables et intelligents. De la sorte, la robotique militaire garantira au soldat une exposition moindre aux dangers qu'il pourrait rencontrer lors de sa progression dans l'espace de bataille. On pense, notamment aux déploiement dans des milieux contaminés, au déminage terrestre et sous-marin ou encore à la lutte contre des essaims de drones.

Enfin, un apport plus controversé des technologies de l'IA au bénéfice du combattant concernerait l'aide qu'elles pourraient fournir pour l'analyse opérationnelle sur le terrain. En rendant possible une identification plus fine des combattants au sein de la population, l'IA contribuerait à un meilleur respect des valeurs du droit international humanitaire. Nous aurons l'opportunité de débattre de ces questions plus loin dans ce travail.

8. Maintenance prédictive

Un autre apport fondamental de l'IA réside dans la « maintenance prédictive ». Les matériels militaires sont souvent soumis à des stress importants et spécifiques, dû à la topologie de la zone d'opération, au climat, aux contraintes environnementales et aux contextes de combat. Or la seule mise en œuvre d'un système d'armes technologiquement complexe suffit également à générer des failles ou des fragilités de structure. La logistique et la maintenance constituent donc des domaines dans lesquels l'IA peut offrir une valeur ajoutée certaine. L'apport de l'IA en matière de maintenance conférera aux armées qui intégreront ce type de technologies un avantage certain sur le plan de la préparation (*readiness*) et du déploiement de forces. Connaître à l'avance les plates-formes, systèmes et logiciels qui exigent un entretien ou une mise à niveau, avant même que les pannes ou les failles n'apparaissent et produisent leurs effets, est devenu un enjeu prioritaire en matière opérationnelle.

Aux États-Unis, La branche innovation de l'armée, la *Defense Innovation Unit* (DIU) a attribué à C3.ai, une société d'intelligence artificielle basée en Californie, un contrat pour un logiciel prédictif capable de déterminer quand les avions militaires auront besoin de réparations. Cet outil rendrait plus d'avions disponibles pour des missions et pourrait potentiellement faire économiser des milliards de dollars en coûts de maintenance à l'U.S. Air Force.

En fin 2017, C3.ai avait conclu un prototype d'accord pour évaluer et traiter les dossiers d'entretien du E-3 Sentry (AWACS) et pour planifier les réparations. L'accord a ensuite été étendu à l'évaluation du C-5 Galaxy et au F-35. Cette période de prototypage s'est terminée en décembre 2019. Pendant cette période, C3.ai est parvenu à prévoir environ un tiers des événements de maintenance non programmée des sous-systèmes.

L'intelligence artificielle pourrait permettre au Pentagone de réaliser une économie de plusieurs milliards de dollars par an, selon certains experts. Dans le rapport annuel 2018 de l'organisation, les responsables ont écrit que, si le DoD suivait le prototype de C3.ai sur toutes les plates-formes d'aéronefs, il pourrait économiser environ 3 à 5 milliards de dollars par an en dépenses de maintenance, grâce à la maintenance prédictive des avions.

Les experts en sécurité nationale s'attendent à ce que le Pentagone adopte plus largement l'intelligence artificielle pour la gestion, avant d'expérimenter des niveaux d'autonomie plus importants sur le champ de bataille. De même, le DIU a fait de la maintenance prédictive l'une de ses principales priorités ces dernières années. Les dirigeants de l'organisation ont également discuté des programmes de maintenance prédictive avec la Marine et l'Armée de terre, notamment pour l'entretien des véhicules terrestres.

En Europe, également, la maintenance prédictive rendue possible par l'IA pourrait s'avérer un axe de développement capital pour les armées. C'est en tous cas l'un des choix qu'a fait la France, selon sa ministre de la Défense, Florence Parly. L'IA au service de la maintenance prédictive portera notamment sur l'entretien des frégates multi-missions et les avions *Rafale*. La France s'est ainsi dotée d'une agence consacrée à l'IA : Agence de l'innovation de défense. Celle-ci qui sera dotée à terme d'un budget de 100 millions d'euros et de compétences en matière d'intelligence artificielle, avec le recrutement rapide de 50 experts dans le domaine. Parmi les projets figure l'intégration de l'IA dans le cadre des besoins de l'avion de combat du futur, notamment pour ce qui a trait à la maintenance prédictive du systèmes d'armes et de ses sous-systèmes[40].

9. Cyberconflictualité et cyberdéfense : quel rôle pour l'IA ?

C'est sans doute dans les domaines de la cybersécurité et de la cyberdéfense que les atouts de l'intelligence artificielle s'exprimeront le mieux. Les enjeux de sécurité liés à l'intégrité des systèmes informatiques qui régissent la quasi-totalité de nos infrastructures sociétales, économiques, administratives et militaires sont réellement colossaux. La vulnérabilité de l'ensemble de ces systèmes, compte tenu de leur étendue et de leur diversité, est réelle. La protection contre cette vulnérabilité ne peut toutefois être assurée par les seuls opérateurs humains derrière leurs terminaux, pas plus qu'elle ne peut être garantie par des systèmes d'antivirus et de pare-feux devenus désuets face à la technicité des méthodes employées et à la furtivité des attaques menées. On peut reconnaître à la cyberattaque des similitudes avec l'arme de destruction massive en ce sens que le rapport entre l'effet physique et l'effet symbolique est très similaire. Dans une certaine mesure, la cyberattaque confère un pouvoir égalisateur à celui qui en use. En effet, tant les capacités que les effets de l'attaque ne sont pas immédiatement corrélés au niveau de puissance militaire[41]. Il importe toutefois d'ajouter que, contrairement à une croyance erronée, les groupes terroristes, même s'ils peuvent disposer pour certains de moyens susceptibles de déstabiliser l'intégrité informatique de services, ne sont pas suffisamment armés pour conduire de vastes opérations de cyberattaques. Dans le domaine des cyberconflits, nous avons davantage affaire à des groupes d'attaquants (ou « hackeurs ») œuvrant au bénéfice de services de renseignement, d'espionnage et de contre-espionnage d'États. Leurs actions visent toute infrastructure critique reposant sur une architecture informatique connectée : compagnies d'eau, d'électricité, de gaz, réseaux commerciaux, administrations publiques, cliniques et hôpitaux, registres et bases de données, aéroports, systèmes ferroviaires, etc. Les systèmes

[40] http://www.opex360.com/2018/03/17/ministere-armees-lance-etude-integrer-lintelligence-artificielle-a-laviation-de-combat-futur.

[41] Un État extrêmement puissance à l'instar des États-Unis peut souffrir des conséquences d'agressions cybernétiques sur ses systèmes, y compris ses systèmes stratégiques.

stratégiques les plus essentiels, tels que ceux qui concernent la dissuasion nucléaire ou la gestion des systèmes d'armes conventionnels (aéronefs de combat, groupes aéronavals, systèmes de missiles et d'antimissiles, liaisons satellites, imagerie spatiale, etc.) ne sont pas à l'abri de telles attaques.

a) Du concept de cyberconflictualité

Si la cyberguerre n'est pas encore une réalité, la cyberconflictualité fait partie du quotidien des relations internationales depuis un certain nombre d'années. Pour comprendre la manière dont l'IA pourrait représenter un atout dans la lutte contre les menaces cybernétiques, encore faut-il comprendre les enjeux de la cyberconflictualité. Il est, en effet, particulièrement difficile de se représenter les formes d'expression de la cyberconflictualité, dans la mesure où ses effets s'entendent moins en termes de « destruction » que de « disruption ». Autrement dit, l'objectif premier des acteurs de la cyberconflictualité est la recherche de la rupture des systèmes plutôt que leur annihilation. Il existe donc une opposition fondamentale entre les effets réels des cyberattaques et la perception de leur dangerosité.

Le cyberespace constitue un espace stratégique complexe. Pour apprécier la nature spécifique de ce cyberespace, sans doute est-il opportun de revenir à des notions plus fondamentales de la stratégie et de comprendre dans quel ensemble conceptuel s'inscrit la notion de « réseau ». Pour Laurent Henninger, le développement contemporain des réseaux, et notamment de l'internet – plus globalement, le cyberespace – constitue l'expression ultime d'un processus de mutation civilisationnelle dont la source peut être historiquement située… au XVI^{e} siècle. À partir de cette période, observe Laurent Henninger, on assiste à un déplacement de l'homme des espaces « solides » vers les espaces « fluides ». L'illustration majeure de cette mutation a lieu au travers des grandes expéditions maritimes et des grandes découvertes. Le rapport de l'homme à la mer traverse alors une transformation majeure : « les étendues marines [deviennent alors] un objectif en tant que tel, […] un démultiplicateur de puissance. »[42] Les espaces fluides peuvent être opposés, sur le plan conceptuel, aux espaces solides. Ces derniers sont de nature « visqueuse » et représentent les seuls lieux où l'homme peut vivre. À l'inverse, indique Laurent Henninger, les espaces fluides sont lisses, isomorphes et inhabitables par l'homme. En conséquence, pour y évoluer, l'homme se doit de recourir à des prothèses techniques, à des traitements mathématiques pour y créer des « réseaux ». Au sein des espaces fluides, le rapport entre le « temps » et l'« espace » sont inversés avec une prédominance très nette du « temps » par rapport à l'« espace ». On a pu constater cette prédominance lors des nombreuses crises qui ont émaillé le début de ce XXI^{e} siècle, et ce dans différents champs de l'action de l'homme. L'espace ne constitue plus un obstacle à l'action humaine à travers les réseaux, il est désormais facilement franchissable, presque secondaire – oserait-on dire négligeable… Qu'il s'agisse des krachs boursiers provoqués par un emballement des intelligences artificielles se situant à la base du trading haute fréquence, de la mobilisation en quelques semaines de moyens militaires considérables pour l'invasion et l'occupation de l'Afghanistan et ensuite de l'Irak, ces quelques exemples n'offrent qu'un aperçu infime de ce à quoi conduira demain la prédominance des réseaux. Il va de soi que l'on ne peut opposer de manière radicale et irréconciliable les espaces fluides et solides. Il existe, comme l'explique Laurent Henninger, des degrés de fluidité et de solidité. Les espaces solides et fluides comportent des interfaces, des lieux d'intersection. Toutefois, il est désormais certain que la maîtrise des espaces fluides et des réseaux constitue une condition *sine qua non* à la domination des espaces solides[43].

[42] Laurent HENNINGER, « espaces fluides et espaces solides : nouvelle réalité stratégique », *Défense nationale*, octobre 2012, numéro 753, p. 1.

[43] Réalité que les Anglo-Saxons semblent avoir intégrée depuis longtemps au travers de leurs marines, aviations, banques, systèmes financiers, médias et réseaux informatiques.

Comment, aujourd'hui, s'expriment les conflits dans les espaces fluides et, plus spécifiquement, les réseaux ? Quelques exemples suffiront à comprendre l'étendue des procédés par lesquels le cyberespace et les réseaux peuvent constituer de nouveaux terrains de rivalité entre puissances. En janvier 2018, lors de la conférence S4 à Miami, les chercheurs de FireEye dévoilaient les secrets du logiciel malveillant Triton qui avait infecté l'un des systèmes de contrôle et d'acquisition de données de Schneider Electrics en Arabie saoudite. Le maliciel opérait en déclenchant des commandes via des appareils industriels et risquait ainsi de mettre des vies en danger. Le 27 juillet 2018, une attaque informatique attribuée au groupe russe APT28 est dévoilée par des chercheurs de Enjoy Safer Technology (ESET). Le procédé a reposé sur la diffusion d'un logiciel malveillant, appelé Lojax, au travers de cibles étatiques situées en Europe centrale et orientale. Immédiatement après le démarrage de Windows, un rootkit UEFI/BIOS déclenchait l'activation d'un second maliciel permettant l'exfiltration de données et l'endommagement des ordinateurs hôtes avant leur propagation[44]. Dans la mesure où Lojax était installé au cœur même du noyau du système, une réinstallation de Windows après formatage ne permettait aucunement la restauration d'un système sain. Le procédé employé par le groupe de hackers russes marquait une véritable innovation dans les modes d'attaque connus. Si la possibilité de recourir aux rootkits UEFI était connue sans que les mécanismes profonds soient réellement maîtrisés, l'attaque par le maliciel Lojax en confirma la fonctionnalité. Ces deux exemples ne sont que d'infimes illustrations de la multitude d'attaques qui, sans discontinuité, s'en prennent aux infrastructures, ordinateurs et réseaux aux quatre coins de la planète, au travers de la vaste étendue du cyberespace.

Si le rôle de l'IA en matière de sécurité cybernétique des systèmes et infrastructures s'avérera déterminant dans les décennies à venir, il convient de comprendre que, en l'état actuel des technologies et des données destinées à alimenter les solutions d'IA, des progrès considérables se doivent d'être accomplis. Aussi paradoxal que cela puisse apparaître, et tout en admettant que le cerveau humain ne pourra à lui seul suffire à contrecarrer les attaques sur les réseaux, les systèmes d'IA actuels se révèlent insuffisamment calibrés pour défendre nos systèmes informatiques en réseaux contre des assaillants cybernétiques et leurs programmes. Les réseaux de neurones profonds actuellement conçus ne sont pas en mesure d'atteindre un niveau de perfectionnement semblable à celui du cerveau humain[45]. À dire vrai, la vaste majorité des systèmes d'IA qui sont aujourd'hui déployés pour la défense des systèmes cybernétiques de nos infrastructures s'appuient sur des architectures de *machine learning* et de *deep learning* dont les algorithmes datent des années 1980. Certes, de tels outils peuvent repérer des groupes, identifier des corrélations, classifier des attaques ou encore détecter des signaux faibles dans les environnements numériques (traces réseaux, systèmes, textes, images, vidéos, etc.), mais ils sont en revanche incapables de détecter des menaces persistantes avancées[46] ou de découvrir l'enchaînement réel des actions qui ont provoqué les dommages constatés. La raison principale de ces insuffisances tient précisément au caractère confidentiel des données piratées. Cette confidentialité appliquée aux données empêche les systèmes

[44] Olivier GESNY, « Capter l'IA de demain au regard des enjeux de cyberdéfense », *Revue Défense Nationale*, dossier « L'intelligence artificielle et ses enjeux pour la Défense », mai 2019.

[45] *Ibid.*, *Revue Défense Nationale*, dossier « L'intelligence artificielle et ses enjeux pour la Défense », mai 2019.

[46] *Advanced Persistent Threats*, selon l'expression créée par l'US Air Force en 2006. Les APT consistent en des attaques informatiques perpétrées sur la longue durée (la première de ce type qui fut détectée en 2003, Titan Rain, s'étala sur trois années) à l'encontre d'États ou de groupes spécifiques représentant d'importants enjeux économiques. L'attaquant est souvent lui-même une organisation de grande envergure, car le degré de discrétion et de mobilisation dans le temps qu'implique une telle attaque suppose d'importants moyens financiers et matériels. Les attaques de type APT reposent sur une stratégie à plusieurs niveaux sur le temps long : organisation spécifique en fonction de la cible visée, développement d'une tactique de recueil de données, couverture des canaux de recueil des données et de dépôt des codes malveillants sur l'infrastructure. L'enjeu principal d'une telle attaque est de garantir le réemploi des chemins empruntés dans le système pour la poursuite ultérieure de l'action de piratage ou de recueil des données.

d'IA de disposer de ressources issues d'attaques réelles pour renforcer leur apprentissage. Il existe donc toujours un décalage entre l'attaque cybernétique et les systèmes sensés contrecarrer cette dernière.

Le recours à l'IA en matière de cybersécurité n'est pas sans présenter quelques vulnérabilités. Compte tenu du développement exponentiel des solutions d'IA, de nouvelles méthodes de piratage et de contournement des processus de contrôle par l'IA ont vu le jour. Cette évolution peut se comprendre. Les technologies d'IA sont aujourd'hui en plein essor et de plus en plus d'applications les mettent en œuvre pour la sécurité des infrastructures informatiques. La connaissance approfondie des protocoles selon lesquels fonctionnent les solutions de sécurité à base d'IA constitue un investissement certes lourd mais très rentable pour toute organisation désireuse de contrer les mécanismes mis en œuvre par de tels dispositifs.

Quelles sont les options qui s'offrent aux attaquants d'une solution d'IA ? Elles sont de plusieurs types.

La première peut consister dans le détournement du fonctionnement même de l'application mettant en œuvre une IA. Il s'agira, dans ce cas, de provoquer volontairement une décision erronée de l'application avec un jeu de données choisi. Par exemple, le contournement d'un système de reconnaissance faciale permettra d'obtenir un accès logique ou physique illégitime et de réaliser un vol de données ou de biens.

Une seconde option peut reposer sur le sabotage du fonctionnement de l'IA. Il s'agira alors d'empêcher ou de perturber le fonctionnement de l'application. Ce type de procédé a plus spécifiquement pour but de ternir la réputation de l'entreprise ayant déployé la solution d'IA pour ses activités. C'est ce qui se produisit pour le fabricant de logiciels Microsoft en 2016. Son application de type *chatbot*, baptisée Microsoft Tay, et dont le but était d'étudier les interactions qu'avaient les jeunes Américains sur les réseaux sociaux, fut inondé pendant toute une nuit de tweets abusifs par un groupe d'utilisateurs malveillants du forum 4chan. En moins de 10 heures, le chatbot avait basculé du comportement d'un adolescent à celui d'un extrémiste.

Une troisième solution exploitée par les attaquants de systèmes d'IA réside dans la *rétroconception* de ces modèles d'IA. L'objectif ici poursuivi n'est pas d'altérer le fonctionnement de ces systèmes, mais bien de « copier » celui-ci afin de revendre les secrets de fabrication et de développement qu'il comporte. Cette action aura pour conséquence d'affecter la valeur même du système d'IA en question et de mettre sur le marché des solutions rivales.

Enfin, une quatrième option peut être le détournement de données utilisées par l'application. Plutôt que de tenter d'attaquer le système d'IA de manière directe, la technique mise en œuvre par les pirates consistera à altérer les bases de données sur lesquelles se fonde l'IA pour s'entraîner et accomplir des tâches.

b) Comment attaquer une IA ?

On peut globalement distinguer trois groupes de méthodes d'attaque d'un système d'IA.

Une première technique consistera à *déplacer le centre de gravité du système d'IA*. Elle vise, plus exactement, à empoisonner ce système. Il s'agira de cibler la phase d'apprentissage automatique du système. L'attaquant cherchera à altérer le comportement de celui-ci dans un sens choisi en affectant la nature et le sens des données utilisées pour l'apprentissage. Toutes les applications à base d'IA qui ont recours à un apprentissage automatique sont susceptibles d'être ciblées par ce type d'attaques. Leurs effets peuvent être des plus redoutables et rendre le système d'IA obsolète, surtout lorsque les bases de données employées pour l'apprentissage automatique sont peu maîtrisées par les concepteurs de la solution d'IA (à l'exemple des bases de données publiques ou fournies par des instances externes aux concepteurs du système d'IA).

Un second groupe d'approches repose sur l'*inférence*. Dans ce cas de figure, un attaquant tentera d'expérimenter successivement diverses requêtes sur l'application pour en étudier l'évolution et le comportement. Le résultat visé est tantôt la récupération de données employées par le système d'IA ou encore la copie de l'architecture de ce système. Déceler de telles méthodes peut s'avérer plus complexe qu'il n'y paraît. En général, les attaquants se déguisent en acheteurs de la solution d'IA pour laquelle ils acceptent de payer la prestation de sécurité. Une fois détenteurs d'une licence d'emploi de la solution de sécurité à base d'IA, ils multiplient les requêtes, récupèrent les informations de sorties associées et utilisent l'ensemble de ces données pour entraîner un modèle identique à la solution de sécurité étudiée. Ce procédé leur permet ensuite d'initier des attaques adverses avec plus de facilité, et surtout d'efficacité. Étant propriétaires de l'algorithme et ayant donc accès à toutes les informations conduisant à la décision de la solution, les attaquants pourront utiliser les exemples adverses générés pour propager des attaques qui ne pourront être détectées.

Vient ensuite le troisième groupe d'attaques, l'*évasion*. Par l'intermédiaire de cette méthode, l'attaquant joue sur les données d'entrée de l'application afin d'obtenir une décision différente de celle normalement attendue par l'application. La méthode consiste à créer l'équivalent d'une illusion d'optique pour l'algorithme. On appelle cette illusion un *bruit* (*noise*). Cette altération se veut la plus furtive possible afin qu'elle ne puisse être détectée et multipliée. L'attaquant cherche donc à détourner le comportement de l'application à base d'IA à son avantage et cible ensuite l'application en production, une fois l'apprentissage terminé. Un exemple souvent évoqué pour illustrer ce type de procédé est la méthode qui fut employée par des chercheurs chinois de la Keen Security Labs qui parvinrent à compromettre le logiciel de pilotage d'une voiture autonome de la société Tesla. En l'occurrence, la Tesla S put être déviée de sa ligne grâce à des « autocollants » collés sur les marquages des voies utilisées par la voiture. Les « attaquants » ont donc donné l'illusion au logiciel de conduite de la Tesla S qu'il s'engageait sur une voie fiable alors que le véhicule évoluait sur une piste non reconnue au préalable. Ce type de faille de sécurité est particulièrement préoccupant dans la mesure où il a concerné, en l'occurrence, un véhicule autonome qui ressemblera demain, très certainement, aux divers systèmes d'armes autonomes qui équiperont les forces armées des pays industrialisés. Si le type de procédé employé par les chercheurs de Keen Security Labs sera certes plus difficile à mettre en œuvre dans le cas de figure d'un déploiement de drones, on peut parfaitement imaginer qu'il puisse néanmoins parvenir à son objectif de détérioration systémique dans le cas du déploiement de robots autonomes lors d'opérations de déminage, de reconnaissance ou de frappes à distance contre des cibles stratégiques.

Un système d'intelligence artificielle ne constitue donc pas la panacée contre les attaques cybernétiques, pas plus qu'il ne représente un bloc.

Catégories d'attaque	Empoisonnement	Inférence	Évasion
Exemple d'attaque	Changer le centre de gravité du système d'IA	Extraction d'informations depuis le système d'IA	Illusions d'optique pour le système d'IA (*noise*)
Motivations	Détournement du fonctionnement/sabotage de l'application	Compréhension et rétroconception du modèle/vol de données	Détournement du fonctionnement
Phase ciblée	Apprentissage	Apprentissage/traitement	Traitement
Facteurs aggravants	Fréquence d'apprentissage/données non maîtrisées	Verbosité des sorties/exposition de l'application à des données d'apprentissage	Complexité des entrants/exposition de l'application

c) La suprématie quantique : une menace pour les systèmes de cryptage ?

Concept des plus contestés par les experts de l'IA et de la physique quantique, la *suprématie quantique* a beaucoup fait parler d'elle depuis 2019. Un rappel des événements – et de leur séquence – est essentiel pour comprendre les enjeux sous-jacents à la prouesse (car il y a bien prouesse, même s'il ne s'agit pas de celle que l'on croit) réalisée par les ingénieurs de Google. Nous détaillerons ensuite les conséquences de cette « avancée » pour le monde militaire, en particulier en matière de sécurité des systèmes d'information.

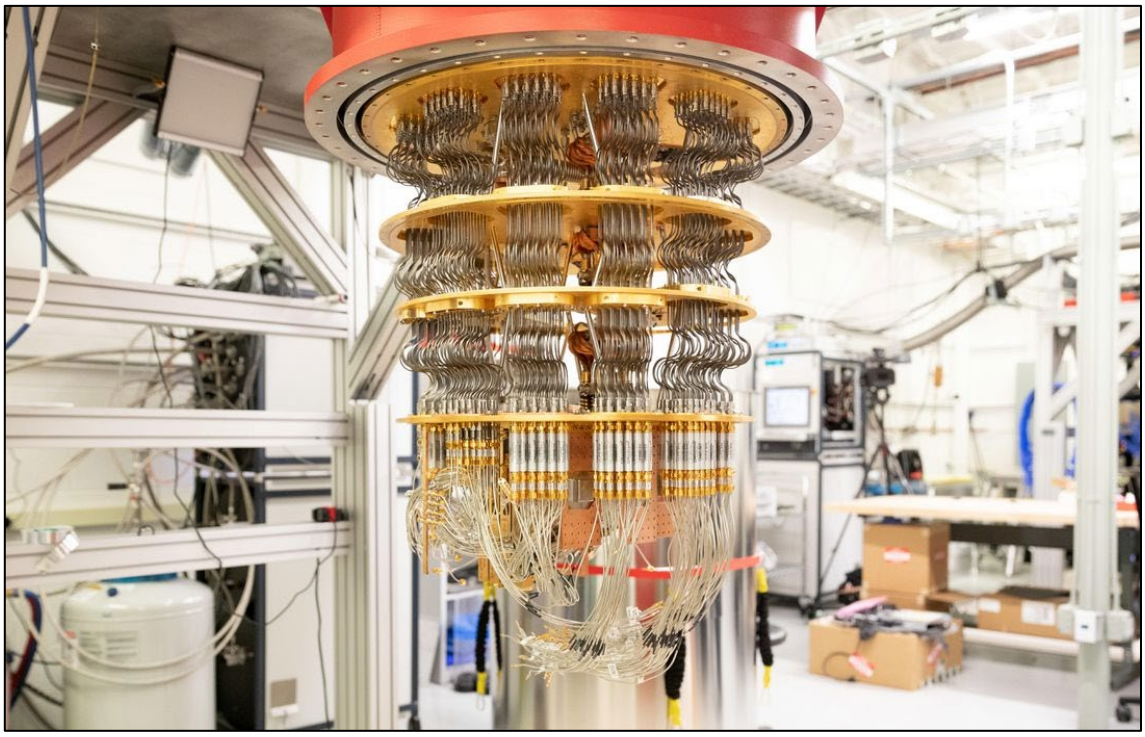


Illustration 3 : l'un des cinq ordinateurs quantiques élaborés par Google, situé dans un de ses laboratoires à Santa Barbara, en Californie

Le 20 septembre 2019, Google publia un article sur le site de la NASA indiquant être parvenu à atteindre la *suprématie quantique*. Dans la foulée de sa mise en ligne, ledit article fut rapidement retiré de la toile, mais déjà de nombreux experts contestaient la portée de l'annonce de Google à travers son ballon d'essai. Un mois plus tard, le 23 octobre, la même équipe de Google à l'origine de l'article de la NASA réitéra son annonce. Cependant, celle-ci prit désormais la forme d'un article dans la célèbre revue *Nature*. L'impression fut différente : on se dit que la rupture annoncée par Google devait être d'une portée sans précédent. Avant de rentrer dans les détails de l'opération réalisée par les ingénieurs de Google, il semble utile d'esquisser rapidement un mot d'explication quant à l'annonce qui fut faite à cette occasion. En faisant état de sa capacité à atteindre le niveau de la suprématie quantique, la société Google voulut signifier qu'elle était désormais capable de concevoir un ordinateur rendant possible la résolution de problèmes inaccessibles aux machines « ordinaires ». En l'occurrence, Google serait parvenue à résoudre en 3 minutes et 20 secondes un calcul qui aurait pris 10 000 ans à l'un des meilleurs supercalculateurs de la planète : le *Summit* d'IBM. À ce stade, deux observations et deux interrogations méritent d'être formulées. La première observation est que Google a une conception pour le moins exotique de la notion de « machine ordinaire ». Cette dernière désigne en réalité un modèle de supercalculateur qui ne procède pas à l'aide de qubits, certes, mais qui diffère néanmoins considérablement des ordinateurs personnels. Une seconde observation consiste à mettre en relief la compétition scientifique et industrielle qui se dissimule derrière cette prouesse, puisque Google

démontre les capacités de son ordinateur quantique en les comparant à celles d'un supercalculateur d'IBM. Cette mise en parallèle n'a rien d'anodin, surtout dans le monde restreint du calcul avancé ! Ensuite, une première interrogation vient à l'esprit : quelle opération a bien pu réaliser l'ordinateur quantique de Google que le meilleur supercalculateur d'IBM n'aurait pu achever dans des conditions de temps similaires ou, à tout le moins, raisonnables ? La seconde interrogation est : la suprématie quantique suppose-t-elle l'avènement d'une intelligence artificielle de « nouvelle génération » ? Nous tenterons de répondre à ces questions avec toute la pédagogie exigée.

i. En quoi a consisté l'expérience de Google ?

Tout d'abord, comment résumer la « prouesse » que Google prétend avoir réalisée en 2019 ? Dans l'étude qu'elle publia dans la revue *Nature*, Google affirme être parvenue à faire travailler son processeur baptisé *Sycamore* pendant 200 secondes pour effectuer un calcul pour lequel le meilleur des supercalculateurs disponibles sur terre – en l'occurrence développé par IBM – aurait mis 10 000 ans. L'ordinateur quantique de 53 qubits de Google aurait ainsi procédé à un calcul dit « spécialisé » qui, dans le cas d'espèce, a consisté à réaliser un échantillonnage de calcul aléatoire en temps polynomial rapide. Que se cache-t-il derrière cette expression ? S'il nous fallait vulgariser l'expérience en question, on pourrait décrire celle-ci de la manière suivante. À l'aide du processeur *Sycamore* qui comporte 53 qubits reliés par des portes logiques, l'ordinateur quantique de Google a lancé une séquence aléatoire de désignation des portes logiques fonctionnelles pour chaque qubit (des 0 et des 1 superposés) à l'aide de la quantique. Une fois que cette opération s'est révélée possible, Google a lancé plusieurs séquences pour finalement évaluer la probabilité d'avoir chacune des 2^{53} combinaisons possibles. En d'autres termes, Google a fait en sorte que l'ordinateur quantique calcule… ce qu'il était en train de calculer. La trivialité apparente de cet ensemble d'opérations ne doit pas nous distraire du caractère fondamental de l'expérience. Lors de cette opération, Google est parvenue à faire fonctionner et à contrôler 53 portes logiques ensemble dans un temps de cohérence raisonnable à ultra-basse température.

Bits classiques	**Bits quantique (qubits)**
2 bits = 2 informations	2 qubits = 4 informations $C_1 \mid 00> + C_2 \mid 01> + C_3 \mid 10 + C_4 \mid 11>$
3 bits = 3 informations	3 qubits = 8 informations $C_1 \mid 000> + C_2 \mid 001> + C_3 \mid 010 + C_4 \mid 011> +$ $C_5 \mid 100> + C_6 \mid 101> + C_7 \mid 110 + C_8 \mid 111>$
4 bits = 4 informations	4 qubits = 16 informations $C_1 \mid 0000> + C_2 \mid 0001> + C_3 \mid 0010 + C_4 \mid 0011> +$ $C_5 \mid 0100> + C_6 \mid 0101> + C_7 \mid 0110 + C_8 \mid 0111> +$ $C_9 \mid 1000> + C_{10} \mid 1001> + C_{11} \mid 1010 + C_{12} \mid 1011> +$ $C_{13} \mid 1100> + C_{14} \mid 1101> + C_{15} \mid 1110 + C_{16} \mid 1111>$
À chaque fois que l'on ajoute 1 qubits, le nombre d'informations est doublé. Le nombre d'informations contenu dans *N* qubits superposés vaut 2^N.	

Pour mieux comprendre encore le résultat de l'expérience menée par Google, il convient de revenir aux fondamentaux. Schématiquement, un qubit peut être considéré comme l'équivalent dans le calcul quantique d'un bit classique. Alors que ce dernier peut prendre comme valeur 0 ou 1, un qubit joue avec les propriétés étranges du monde quantique et consiste en une *superposition* de 0 et 1. Un qubit peut dès lors représenter deux états, deux qubits correspondre à quatre états, etc. *N* qubits équivaut à 2^N informations et 70 bits quantiques contiennent 2^{70} informations (soit 1.000.000.000.000.000.000.000 ou 1 zettabit d'informations). Selon les experts, il s'agirait de la quantité d'informations produite par l'ensemble de l'Humanité depuis son origine !

Pour l'ordinateur quantique de Google, les 53 qubits équivalent à plus de 9×10^{15} états superposés de 0 et 1. En plus de la *superposition*, une autre propriété essentielle des ordinateurs quantiques est l'*intrication*. L'*intrication* désigne plus exactement un phénomène par lequel les états quantiques de deux entités (ou plus, à l'instar des qubits) sont dépendants les uns des autres. L'intrication et la superposition permettent un parallélisme massif dans le traitement de l'information. Ces deux propriétés constituent la base même de la puissance des ordinateurs quantiques. Toutefois, un problème survient lorsque l'on fait fonctionner un système quantique : la *décohérence*. En effet, les interactions entre le système quantique et l'environnement (ondes électromagnétiques, rayonnements, etc.) conduisent à une perte des propriétés du système quantique. Et cette décohérence est d'autant plus importante que le nombre de qubits est élevé. Tout l'enjeu de la computation quantique – et donc, de la course engagée entre les détenteurs de systèmes de calcul quantique – consiste ainsi à maintenir le plus longtemps possible un système quantique dans un état de fonctionnement à très basse température (20 millikelvins soit une température à proximité du zéro absolu) en isolant celui-ci des perturbations engendrées par l'environnement qui pourraient déstabiliser l'équilibre dudit système. En réalité, un ordinateur quantique est essentiellement une expérience de laboratoire contenue dans un cryostat.

Actuellement, sur le plan pratique, l'ordinateur quantique ne présente aucune utilité. Il est d'ailleurs quasi certain que celui-ci ne remplacera jamais l'ordinateur classique pour une immense majorité de tâches. La connaissance et la maîtrise de la computation quantique recèle encore bien des mystères. Et les applications potentielles de cette forme nouvelle de calcul – si elle devait être contrôlée et développée – restent pour la plupart méconnues en l'état actuel de la science et de la technologie. L'une des problématiques auxquelles sont confrontées les équipes travaillant sur la computation quantique a notamment trait au nombre d'erreurs susceptibles de se produire lors d'un calcul quantique. Au risque de caricaturer, plus le nombre d'informations traitées et produites est grand, plus le risque d'erreurs est important. Pour parer à cette contrainte, les processeurs quantiques nécessitent des codes correcteurs d'erreurs. Les *Large Scale Quantum Computers* permettent de pallier les erreurs liées à la décohérence. Cependant, une fois de plus, une contrainte de taille existe, car il faudrait vérifier intégralement et efficacement chaque qubit utilisé pour un calcul. En effet, pour vérifier l'intégrité de chaque qubit de calcul, il faudrait consacrer 1.000 ($2^{1.000}$), voire 10.000 ($2^{10.000}$) qubits physiques spécifiquement au contrôle. Autrement dit, il s'agit là d'un nombre de qubits absolument colossal et difficilement à portée de l'entendement humain.

ii. <u>L'ordinateur quantique peut-il conduire à une nouvelle IA ?</u>

En l'état actuel de la technologie et des connaissances sur la computation quantique, la réponse est non. Si une grande partie de la presse généraliste et spécialisée a pu laisser s'établir un amalgame entre l'ordinateur quantique et l'intelligence artificielle, il importe de différencier ces deux domaines technologiques. Certes, un lien pourrait exister à l'avenir entre ces deux technologies, mais aujourd'hui rien ne prédit avec certitude la variété ou le nombre d'applications susceptibles de découler du modèle d'expérience menée par les équipes de Google. Atteindre la *suprématie quantique* consiste pour un ordinateur – quantique – à pouvoir effectuer n'importe quel calcul impossible à réaliser avec un

ordinateur « classique » (c'est-à-dire un supercalculateur). Une telle « définition » se fonde donc sur une distinction pour le moins artificielle et susceptible de fluctuer avec le temps et les avancées de la science et de la technologie. Le stade de la suprématie quantique dépend tout autant des capacités de l'ordinateur quantique que de l'évolution des supercalculateurs. La question des perspectives d'emploi de l'ordinateur quantique est, à dire vrai, secondaire. Dans les années 1990, les premiers algorithmes quantiques avaient permis la résolution de problèmes artificiels dont personne ne pensait qu'ils étaient vraiment intéressants. Mais les informaticiens avaient pu malgré tout appliquer au développement d'algorithmes ultérieurs les enseignements tirés de leurs expérimentations quantiques. En d'autres termes, des problèmes pratiques avaient pu être résolus grâce au ruissellement des enseignements de l'informatique quantique vers l'informatique classique.

L'expérimentation dirigée par Google a seulement permis de « frapper à la porte » de la suprématie quantique. De nombreuses entreprises à l'instar d'IBM, IonQ, Rigetti ou l'Université de Harvard contestent la portée scientifique de l'annonce faite par Google. Les équipes de ces entités prétendent que d'autres voies de développement d'un ordinateur quantique sont possibles. Google a fait le choix des supraconducteurs qui présentent l'avantage d'être constitués d'un matériau solide et construits avec des techniques de fabrication existantes. Par ailleurs, leurs portes logiques effectuent des opérations très rapides. Néanmoins, de tels circuits doivent être refroidis à des températures extrêmement basses et chaque qubit d'une puce supraconductrice doit être calibré individuellement. Or, cette technique de calibrage peut très vite s'avérer inadaptée pour des milliers de qubits (le nombre record de qubits opérationnels est, pour rappel, de 53). IonQ travaille pour sa part sur d'autres pistes technologiques que sont les ions piégés par laser. Toutefois, la technique des ions piégés offre un bilan plus que contrasté. Si les ions ont l'avantage d'être tous identiques et de présenter, en étant piégés, un temps de stabilité plus long pour effectuer le calcul avant que ne survienne la décohérence, les portes logiques employées par les ions s'avèrent plus lentes (des milliers de fois plus longues que les portes supraconductrices comme celles auxquelles Google recourt avec le *Sycamore*). En conclusion, l'informatique quantique nécessite aujourd'hui un bond technologique majeur pour fonctionner. Pour Adam Bouland, « l'informatique quantique aurait besoin d'une invention analogue au transistor – une technologie révolutionnaire qui fonctionne presque sans faille et qui soit facile à déployer à grande échelle »[47].

Et si nous imaginions un instant que la suprématie quantique était atteinte et démontrée, quelles applications pourraient être tirées de ce saut considérable ? C'est à ce stade de la réflexion que les conjectures se multiplient, se confondent et, parfois, se dissipent. On appelle « avantage quantique » le fait de tirer de la suprématie quantique des applications concrètes. Pour l'heure, les experts évoquent les secteurs d'activités dans lesquels l'ordinateur quantique pourrait révéler une certaine utilité : services financiers, chimie et intelligence artificielle. De nombreuses entreprises attendent avec impatience les solutions qui seront offertes demain par l'informatique quantique. Le calcul quantique devrait permettre à l'avenir de créer des modèles de villes et même de galaxies, d'optimiser les flux de circulation, de modéliser des molécules d'ADN et des matériaux innovants ou de formuler des pronostics financiers[48]. Il est toutefois un domaine dans lequel des attentes particulières sont nourries : le cryptage. Les perspectives d'applications issues du calcul quantique pour le secteur des communications et des réseaux sont nombreuses : protection des réseaux de communication et d'information, sécurité des données bancaires, protection des institutions et services publics, etc.

[47] Kevin HARNETT, « Suprématie quantique : le guide pratique », *Pour la Science*, Hors-Série, mai-juin 2020, pp. 86-90.

[48] Zaour MAMEDIAROV, « Bataille mondiale pour la suprématie quantique », *Courrier international*, *cf.* https://www.courrierinternational.com/article/technologies-bataille-mondiale-pour-la-suprematie-quantique.

On pense encore au développement de capteurs quantiques dans plusieurs domaines tels que la sécurité et la défense, la navigation (notamment dans le secteur spatial), etc.

Le calcul quantique offrirait également des perspectives nouvelles pour l'internet du futur. La *Quantum Internet Alliance* (QIA) est une initiative européenne visant à élaborer un réseau de transmission d'information quantique qui viendrait s'ajouter – et non se substituer – à l'actuel internet. Il s'agirait, en quelque sorte, d'ajouter des fonctionnalités nouvelles et spéciales au réseau global, notamment des fonctionnalités relatives à la sécurisation des communications. Les équipes du QIA œuvrent à l'élaboration de clés quantiques pour la protection des communications. Leur capacité de résistance proviendrait des vertus de l'*intrication*. À titre d'exemple, un qubit situé à Paris et un qubit situé à New York pourraient être intriqués grâce à un internet quantique. Cela signifie que la mesure de chacun des deux qubits en question aboutirait toujours à la même et unique mesure. Cette perspective ouvre la voie à une coordination sécurisée à distance. Par ailleurs, l'avantage de l'intrication quantique est d'être *privée*. Dès lors que deux qubits distants sont intriqués, rien d'autre ne peut partager leur information. Cette piste de recherche pourrait aboutir à de nouvelles formes d'informatique à distance.

Le calcul quantique n'aboutira pas en soi à une nouvelle forme d'intelligence artificielle. Il s'avérera néanmoins un adjuvant fondamental pour l'augmentation des puissances de calcul sitôt que les concepts, les matériaux et les technologies permettront de stabiliser les applications quantiques qui n'en sont, pour l'heure, qu'au stade expérimental. Le calcul quantique et le qubit viendront sans doute se superposer aux méthodes de calcul avancé contemporains, sans toutefois les remplacer intégralement.

d) La dissuasion à l'ère de l'IA et de l'infosphère

L'avènement de l'ère des ordinateurs et des réseaux a fait émerger avec elle de nouvelles interrogations sur les périls liés à la dépendance toujours plus grande des nations les plus avancées aux technologies de l'information. L'interconnectivité croissante des systèmes de défense et l'interpénétration des ordinateurs et des armements ont fait surgir des craintes nouvelles quant à l'existence de failles informatiques pouvant être exploitées par des adversaires. C'est surtout aux États-Unis que les craintes les plus marquées se sont fait jour à l'égard de ce qui pourtant, au premier abord, semblait représenter un vecteur supplémentaire de domination des forces armées américaines en plus de celle détenue dans les domaines conventionnel et nucléaire. La dépendance des États-Unis et d'autres nations avancées aux systèmes informationnels s'entend sous deux aspects. Le premier est la porosité des infrastructures sociétales (administrations, services publics, écoles, hôpitaux, régies des eaux, du gaz, de l'électricité, centrales nucléaires, etc.) aux menaces nouvelles susceptibles d'exploiter les fragilités des architectures informatiques. Les attaques frappant les instances relevant de cette catégorie d'acteurs sont regroupées sous le terme de *netwar* (guerre de réseau). Le second aspect de la dépendance informationnelle concerne plus précisément les installations militaires et les instances décisionnelles stratégiques. Toute forme d'attaque s'en prenant expressément à ces institutions relève de la *cyberwar* (ou guerre cybernétique). D'une certaine façon, nous retrouvons là des similitudes avec les fondements de la dissuasion nucléaire et la distinction essentielle entre la stratégie contre cités et la stratégie contre forces. Pourtant, les ponts conceptuels qui peuvent exister entre la guerre à l'âge des réseaux et de l'IA, d'une part, et les principes de la dissuasion nucléaire, d'autre part, comportent des limites. De nombreux experts admettent que l'âge des réseaux et l'émergence de l'IA feront apparaître des enjeux nouveaux en matière de sécurité.

10. Vulnérabilités potentielles

a) Le risque d'insuffisance de données d'alimentation

L'intégration de l'IA dans le domaine militaire présente en outre des défis techniques communs à l'ensemble des utilisateurs de cette technologie. La plupart concernent les données disponibles pour l'IA, la quantité et la qualité de ces données étant les principaux « ingrédients » pour de bons algorithmes. Il existe à cet égard deux défis de taille.

Les données d'entrée sont un élément central des algorithmes de l'IA. La qualité d'un algorithme dépend, d'une part, des données d'entraînement utilisées avant son intégration dans un produit ou un service et, d'autre part, des données récupérées lors de l'utilisation de l'algorithme dans le monde réel. Cela conduit à ce que l'on appelle la « vulnérabilité de l'alimentation en données »[49].

Dans la mesure où il est particulièrement difficile de recueillir des ensembles de données qui soient suffisamment volumineux et représentatifs des situations du monde réel, l'IA reproduit les distorsions présentes dans ses données d'entraînement[50]. Ainsi, l'algorithme GloVe, qui associe des mots présentant une similitude sémantique, a été formé à l'aide de 840 milliards d'exemples tirés du web : le constat est qu'il avait fortement tendance à reproduire les préjugés sexistes et racistes[51].

Même bien entraînés, les algorithmes peuvent être très instables. Un aspect très important – peut-être plus que les autres – est que les systèmes intégrant l'IA sont incapables de s'adapter, ou s'adaptent mal, à de nouveaux contextes, même s'ils ont un fonctionnement très similaire au cerveau humain[52]. À titre d'exemple, un algorithme entraîné exclusivement à l'aide de sources formelles n'est pas capable de comprendre des contenus informels tirés des réseaux sociaux. Pour citer un autre exemple, un système intégrant l'IA a correctement conclu, avec 57,7 % de confiance, qu'une image de panda représentait effectivement un panda. Lorsque l'image a été modifiée de seulement 0,04 % sur sa qualité en pixels, le système soudain été convaincu à 99,3 % que l'image représentait un gibbon. On constate donc qu'une modification minime peut induire le système en erreur totale[53].

La vulnérabilité de l'alimentation en données est sensiblement exacerbée dans le secteur de la défense. Les données y sont souvent très rares, en comparaison avec le secteur civil. Ainsi, les données dont disposent les forces aériennes sur le comportement de leur aéronef pendant des opérations de combat sont dérisoires par rapport à la réserve de données auxquelles ont accès les compagnies aériennes privées. De surcroît, le personnel militaire doit souvent opérer dans des environnements où les données sont extrêmement réduites et les situations très incertaines, par exemple dans des contextes difficiles comme l'Afghanistan[54].

b) La fiabilité

Si le personnel militaire est invité à adopter des systèmes intégrant l'IA, il doit pouvoir avoir l'assurance qu'ils fonctionneront comme prévu. Or, ces systèmes continuent de connaître de sérieux problèmes

[49] Osonde OSOBA, William WELSER IV, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*, Santa Monica (Calif.), RAND Corporation, 2017.
[50] *Ibid.*, Santa Monica (Calif.), RAND Corporation, 2017.
[51] Jean-Christophe NOËL, *Intelligence artificielle : vers une nouvelle révolution militaire ?*, Paris, Institut français des relations internationales (IFRI), Laboratoire d'études stratégiques, coll. Focus stratégique, 2018.
[52] Daniel S. HOADLEY, Nathan J. LUCAS, *Artificial Intelligence and National Security*, Washington, Congressional Research Service, 7-5700, R45178, April 26, 2018.
[53] Jean-Christophe NOËL, *op. cit.*, 2018.
[54] Lindsey R. SHEPPARD & al., *Artificial Intelligence and National Security: the Importance of the Ecosystem*, Washington, Center for Strategic and International Studies, November 2018.

de fiabilité. En règle générale, le niveau de confiance doit être beaucoup plus élevé dans le secteur de la défense que dans le domaine civil. Lorsqu'un site de vente en ligne recommande des produits qui n'intéressent pas le consommateur, il n'y a pas grand mal à cela. En revanche, lorsqu'un système militaire intégrant l'IA fait des erreurs, les conséquences peuvent être beaucoup plus graves, et peuvent même aller jusqu'à la perte de vies humaines.

À l'heure actuelle, il est encore très difficile – et parfois impossible – de déterminer si les systèmes faisant appel à l'IA tirent les bonnes conclusions, voire *comment* ils les tirent. Ces systèmes apparaissent souvent comme des « boîtes noires » aux yeux des chercheurs et des opérateurs. Il arrive que les algorithmes produisent des résultats « étranges », résolvent les problèmes en utilisant une méthode fausse ou contraire à la logique, ou même « trichent »[55]. La garantie d'une IA « explicable » ainsi que la nécessité de mettre en place des processus de validation et de vérification spécifiques à cette technologie est donc devenue indispensable.

Dans la mesure où les systèmes faisant appel à l'IA sont très dépendants de l'exactitude des données, ils sont aussi très vulnérables à la manipulation des données d'entrée, notamment lors de cyberopérations. Bien que le volume des données traitées soit souvent élevé, la modification même minime d'un algorithme peut avoir des effets catastrophiques. Dans le domaine de la classification des images, par exemple, il a été prouvé que la modification d'un seul pixel pouvait suffire à tromper l'algorithme[56]. Dans une étude récente, un algorithme de classification d'images a identifié de façon erronée une mitrailleuse comme étant un hélicoptère[57]. Un autre angle d'attaque est celui des données d'entraînement. Les réseaux neuronaux profonds utilisent souvent des modèles s'appuyant sur des données d'entraînement provenant de tierces parties. Ces données pourraient donc être une cible intéressante pour les adversaires. Les systèmes intégrant l'IA peuvent être eux-mêmes la cible d'attaques : des acteurs mal intentionnés peuvent essayer de voler ou de dupliquer leur contenu, que ce soit pour l'intégrer dans leurs propres systèmes ou pour trouver des moyens de neutraliser ces systèmes.

Aujourd'hui, le déploiement de l'intelligence artificielle, bien qu'en progression constante, se situe encore au stade du balbutiement. Et les cas où cette technologie est employée, la tolérance aux erreurs s'avère encore grande. Dès lors qu'une industrialisation des technologies de l'IA serait envisagée, des exigences plus élevées, notamment à l'endroit de ses capacités de résistance, apparaîtraient. Actuellement, et malgré les évolutions notables dont la technologie a bénéficié, les réseaux de neurones profonds restent encore victimes de manipulations.

De la même façon, les techniques d'apprentissage d'IA présentent-elles de nombreux risques. On citera, par exemple :

- les biais involontaires survenant en présence de données d'apprentissage non représentatives (par exemple, biais ethnique dans des données de population) ;
- les biais volontaires provenant de la modification intentionnelle des données d'apprentissage ou du modèle de traitement de ces données ;
- les failles intervenant dans la reconstitution de données d'apprentissage particulièrement sensibles ;
- les cas de résultats opaques ou peu explicables.

[55] *Ibid*., Washington, Center for Strategic and International Studies, November 2018.
[56] Peter SVENMARCK & al., "Possibilities and Challenges for Artificial Intelligence in Military Applications", in *Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting*, NATO Science and Technology Organisation, 2018.
[57] Daniel S. HOADLEY, Nathan J. LUCAS, *op. cit.*, 2018.

La qualité des données d'apprentissage constitue un prérequis fondamental à la conception d'algorithmes robustes et de systèmes d'IA fiables. Toute altération, faille corruption de ces données, voire même leur absence en quantité suffisante, conduirait à l'élaboration de systèmes d'IA inadaptés et erronés.

Il faut enfin mentionner le risque auquel pourrait conduire une dépendance trop grande de nos sociétés, entreprises et organisations à une technologie qui viendrait à simplifier de façon exagérée certains outils, au point que les agents humains ne disposeraient plus des compétences nécessaires tant pour la compréhension que la conduite de leur mission. Certains risques liés à l'apprentissage des systèmes d'IA pourraient, en effet, ne pas être détectés. Ce serait notamment le cas si un système d'IA était victime de leurres, de portes dérobées ou de rétroconceptions. La mise en œuvre de telles IA corrompue compromettrait le déploiement des solutions supervisées par les organisations qui y auraient recours. Des compétences humaines devraient dès lors être maintenues pour résoudre de telles failles.

11. L'IA, la prochaine bulle spéculative ?

Nonobstant l'ensemble des applications multisectorielles que les promoteurs de l'IA nous promettent, des interrogations légitimes demeurent quant à la réalité de ses potentialités, y compris dans le domaine militaire. C'est une hypothèse qu'il importe de ne pas écarter d'un revers de la main : affirmer l'existence d'une bulle spéculative autour de l'IA ne consiste pas à dénigrer cette technologie et ses perspectives applicatives. Il s'agit, plus exactement, d'indiquer qu'un degré d'attente trop élevé sur les résultats de l'IA peut conduire – et conduira certainement – à une déception massive des investisseurs. Le Dr Laurent Alexandre évoque la problématique de la bulle spéculative de l'IA de la manière la plus lucide qui soit. C'est une véritable ruée vers l'or qui semble aujourd'hui caractériser le domaine de l'intelligence artificielle. En effet, nombre de réussites (et de semi-échecs) ont poussé des investisseurs de divers horizons à soutenir financièrement plusieurs entreprises – parmi lesquelles de nombreuses start-ups (appelées aussi *licornes* en français) – dédiant leurs activités et leur cœur de métier à des applications autour de l'IA. Parmi ces start-ups, il convient de reconnaître qu'un grand nombre ne se consacrent que peu ou prou à des activités touchant de près ou de loin l'IA : le qualificatif « IA » n'a d'autre but que d'attirer les investisseurs les plus aventureux. Pour le Dr Laurent Alexandre, de telles bulles spéculatives sont indispensables à la génération de ruptures technologiques. Certes, elles supposent des niveaux de perte parfois importants pour de nombreux investisseurs, mais ce serait là le prix à payer selon l'auteur[58]. On retrouve dans cet argument le principe de « destruction créatrice » de l'économiste Joseph Schumpeter. Le cœur de son argumentation – qui est aussi celle du Dr Laurent Alexandre – est d'établir une distinction aussi nette que possible entre les concepts de « croissance économique » et de « développement ». Christian Deblock explicite la pensée de l'économiste : « il y a développement [selon Schumpeter] lorsqu'il y a passage, et par le fait même rupture, d'un état d'équilibre à un nouvel état d'équilibre qui n'a rien à voir avec le précédent. À chaque équilibre du système est associée une combinaison spécifique de facteurs de production ; c'est un tout stable que vient bouleverser l'innovation. Le développement est discontinuité, turbulence, et il n'y a développement que lorsqu'il y a " destruction créatrice ", autrement dit une réorganisation du système sous l'effet d'une recombinaison de l'appareil productif ou innovation. »[59] La bulle spéculative, serait-on tentés de dire aujourd'hui, est précisément le phénomène émergent,

[58] Laurent ALEXANDRE, *La guerre des intelligences : comment l'intelligence artificielle va révolutionner l'éducation*, Paris, Seuil, coll. Livre de Poche, 2017, p. 51.

[59] Christian DEBLOCK, « Introduction : innovation et développement chez Schumpeter », *Revue interventions économiques*, volume 46, 2012, p. 46.

annonciateur, d'une transformation des conditions dans lesquelles opère un système de production sur le point d'accoucher d'une rupture.

Craindre la survenance d'une bulle spéculative autour des technologies de l'IA n'altère donc en rien la force de transformation de ces technologies pour l'avenir. Bien au contraire : l'éclatement d'une bulle serait même la condition d'évolution des technologies de l'IA. Un tel événement participerait à un effacement des acteurs qui comptent peu dans le secteur au profit des entités les plus avancées sur le plan de la recherche et du développement.

Sur le plan géopolitique, l'explosion/implosion d'une bulle spéculative dans le domaine de l'IA ne serait pas sans conséquences sur les équilibres des forces entre puissances technologiques. Parmi les craintes manifestées par les États-Unis à l'aube de la « Revolution in Military Affairs » des années 1980 et 1990, figurait l'hypothèse selon laquelle les ruptures technologiques sur lesquelles les forces des États-Unis avaient bâti leur supériorité stratégique pourraient être récupérées et exploitées au maximum par des nations (ou des acteurs) qui ne se situent pas à l'origine de l'innovation. Longtemps, le Japon avait figuré comme le principal rival technologique des États-Unis dans le secteur des technologies des ordinateurs et des réseaux. Ce fut là une erreur, qui s'explique somme toute en raison de l'incapacité récurrente dont ont eu à souffrir les puissances établies à identifier leurs futurs rivaux. Il est en outre fréquent, dans le domaine des relations internationales, d'observer que ce sont les communautés politiques les moins propices à produire de la puissance politique et technique qui parviennent à étendre leur domination sur le monde connu. Les exemples sont nombreux : Athènes, Sparte, Rome, et, plus tard, l'ensemble de l'Europe occidentale. C'est plus particulièrement à propos de cette dernière que l'historien Paul Kennedy écrivait : « au début du XVI^{e} siècle, rien ne laissait présager que cette dernière [l'Europe de l'Ouest] allait s'élever au-dessus du reste du monde. »[60] Le manque de perspectives attribuées rétrospectivement par les historiens au Vieux Continent au XVI^{e} siècle provient souvent d'une survalorisation des vicissitudes auxquelles les nations européennes étaient en proie alors même que les faiblesses perçues s'avéraient au contraire des axes de force pour le développement politique et technologique. À l'inverse, nombre de nations qui auraient pu s'ériger comme des candidats potentiels pour constituer les berceaux d'innovations technologiques majeures virent leurs chances amenuisées, voire effacées, du fait de contraintes internes sociologiques, démographiques ou économiques. Le Japon fait figure d'exemple. Pays vieillissant et déclinant, la force économique du Japon – qui le maintiendra encore pour un temps indéterminé parmi les cinq puissances technologiques de la planète – s'amenuisera du fait de sa situation géographique, de l'encerclement dont il est victime (une Chine compétitive, une République nord-coréenne soumise aux caprices d'un pouvoir dictatorial décadent, etc.). Surtout, le Japon a échoué à fixer sur son sol une communauté de scientifiques étrangers ayant œuvré dans le cadre d'échanges académiques. Cette politique d'ouverture « contrariée » handicape le pays, à l'inverse des États-Unis dont le terreau scientifique et technologique séduit les esprits les plus innovateurs de multiples régions du monde.

Qu'en conclure ? Ainsi que l'observait déjà Richard O'Hundley dans un ouvrage phare de la RAND Corporation, en 1999, à propos du concept de « Revolution in Military Affairs » (RMA)[61], rares sont les révolutions dans les affaires militaires qui furent portées par les puissances dominantes. Il poursuivait : les RMA les plus abouties sont souvent conduites par des acteurs autres que ceux qui se situent à l'origine des innovations technologiques et doctrinales qui les constituent. Il ajoutait encore que les véritables ruptures technologiques à portée militaire impliquaient, outre une ou plusieurs innovations

[60] Paul KENNEDY, *Naissance et déclin des grandes puissances : transformations économiques et conflits militaires entre 1500 et 2000*, traduction de Marie-Aude Cochez et Jean-Louis Le brave, Paris, Payot, coll. Petite bibliothèque Payot, p. 18.
[61] Richard O'HUNDLEY, *Past Revolutions, Future Transformations: What Can the History of Revolutions in Military Affairs Tell Us About Transforming the U.S. Military?*, Santa Monica (Calif.), RAND Corporation, 1999, p. XIV.

technologiques, des organisations adaptées et la mise au point de doctrines à même de déployer les effets de l'innovation.

12. Conclusions partielles

Comme nombre de technologies innovantes dont les résultats pourraient se révéler d'une ampleur disruptive, il est particulièrement difficile d'identifier à l'avance la diversité des applications auxquelles l'IA pourrait aboutir. Comme nous avons pu le constater, les quelques pistes de déploiement de l'IA dans le secteur militaire s'inscrivent dans un large éventail de domaines allant de la gestion automatisée de systèmes en auto-organisation à la personnalisation poussée à l'extrême des équipements du fantassin ou à la production de mondes virtuels toujours plus proches de la réalité du combat pour les besoins de l'entraînement du soldat. En l'état actuel de nos réflexions, il pourrait être dit que, si l'IA ne semble pas donner naissance à des nouveautés jusque-là inexistantes en matière d'applications, elle se révélera néanmoins un formidable multiplicateur d'effets et d'efficacité dans l'éventail des dispositifs qui soutiendront le militaire dans ses missions. Néanmoins, l'adjonction de l'IA au sein des solutions développées au service du soldat laissera apparaître aussi des failles nouvelles qui pourraient, demain, être exploitées par des adversaires qui seront parvenus à maîtriser cette même technologie. L'IA ne modifiera donc en rien ce qu'Edward Luttwak a pu désigner par la logique dialectique de la stratégie. Le prochain chapitre nous permettra d'explorer certains aspects de cette dialectique.

III. Hypothèses et scénarios spécifiques

1. IA et guerre préemptive

L'impact de l'IA sur les équilibres militaires conventionnels et non conventionnels – en particulier, nucléaire – a donné lieu à une considérable production littéraire émanant de *think tanks*, groupes et centres de recherche. Comme nous aurons l'occasion de l'exposer, de nombreux scénarios ont été élaborés à propos des avantages que pourrait apporter l'IA au sein des organisations militaires et, plus précisément, en matière de conduite des opérations militaires. Nous avons déjà eu l'opportunité de détailler les apports possibles et réels de l'IA pour les forces armées. La présente partie sera, quant à elle, consacrée à la transformation des équilibres militaires découlant de la maîtrise des technologies d'IA par un ou plusieurs protagonistes d'un rapport de forces donné. Avant d'explorer les scénarios supposant la possession effective par un acteur stratégique de systèmes d'IA, il convient de nous interroger sur les risques que pourraient faire peser sur la sécurité internationale les entreprises conduites par certains États en vue de parvenir au contrôle des technologies constitutives de l'IA et, en particulier, d'une intelligence artificielle générale (IAG), aussi appelée « IA forte ». Pour rappel, une intelligence artificielle générale ou forte est une IA dont les capacités ne seraient pas limitées à un ou plusieurs domaines mais apte à effectuer tout type d'inférence, d'analyse et de décision dans l'ensemble des champs d'action. L'IAG suppose, comme l'expriment fort bien quelques spécialistes de la question, l'équivalent d'une « explosion cambrienne »[62] dans le domaine de l'intelligence synthétique[63]. Certains n'hésitent d'ailleurs pas à affirmer que l'IAG constituerait, dans une certaine mesure, l'ultime création de notre civilisation technologique. Après elle, aucune autre production matérielle de l'homme ne pourrait venir égaler l'avènement d'une superintelligence. Face à cette perspective, quelle serait l'attitude des États dans l'hypothèse où l'un d'eux approcherait le seuil de l'émergence d'une superintelligence ? Telle est en substance la question qui préoccupe un certain nombre d'analystes. Avant que nous tentions de répondre à cette question, rappelons qu'en l'état actuel de la technologie, il n'existe pas d'IA du type « superintelligence ». Les seuls systèmes d'IA qui existent sont principalement des machines fondées sur l'autoapprentissage dans des domaines délimités. Par ailleurs, une grande part de la communauté scientifique doute qu'il puisse être possible d'assister à l'apparition – ou au développement – d'un telle superintelligence. On peut, il vrai, arguer du fait que l'hypothèse de l'émergence d'une telle IAG appelle plus d'interrogations que de réponses. La première consiste à poser la question du critère qu'il conviendrait de prendre en considération pour affirmer qu'une superintelligence aura été développée. Plus encore, qui ou quelle autorité viendrait à attester de l'avènement d'une telle prouesse technologique ? Du reste, que signifierait-elle ? Comment se traduirait-elle dans les faits ? Aurait-on affaire à un saut quantitatif ? S'agirait-il simplement d'une IA capable de résoudre en quelques minutes des problèmes complexes pour lesquels les meilleurs ordinateurs jusqu'alors auraient pris plusieurs milliers d'années[64] ? Ou aurions-nous affaire à une rupture qualitative avec une IA dotée de conscience comme le laissent suggérer les propos de Ray Kurzweil ? Comme nous pouvons nous en rendre compte, l'approximation par quelque acteur technologique (qu'il représente un État ou une société privée) reste une affaire de perception.

---

[62] L'explosion cambrienne désigne l'âge géologique durant lequel sont apparus soudainement la plupart des grands embranchements actuels d'animaux pluricellulaires. Cette période est caractérisée par une grande diversification des espèces animales, végétales et bactériennes. Certains auteurs évoquent même cet âge comme un « big bang zoologique du cambrien ». Armand DE RICQLÈS, « Un big-bang zoologique au cambrien ? », *La Recherche*, n° 240, février 1992, pp. 224-227.

[63] Anand RAMAMOORTHY, Roman YAMPOLSKIY, "Beyond MAD? The race for Artificial General Intelligence", *ITU Journal: ICT Dicoveries*, Special Issue No. 1, 2 février 2018.

[64] Il est question ici, plus précisément, d'ordinateur quantique.

Pour beaucoup d'analystes, la crainte de voir un État maîtriser une technologie aussi « disruptive » que l'intelligence artificielle générale serait de nature à déstabiliser les équilibres politico-militaires et aboutir à l'irruption d'une guerre préemptive, qui serait menée par un État ou une coalition d'États résolus à empêcher la possibilité qu'un ou plusieurs de leurs homologues ne disposent d'une telle avance technologique. Quelles seraient les raisons sous-jacentes à une telle entreprise ? Un État qui serait sur le point de développer une IAG disposerait d'une allonge considérable sur ses partenaires et adversaires. Avec une telle IAG, la possibilité qu'une autre superintelligence puisse voir le jour ailleurs qu'au sein de l'État pionnier aurait un taux de probabilité proche de zéro. Un État assisté par une IAG – que d'aucuns désignent par l'expression d'« oracle cybernétique » – serait en mesure de calculer avec précision et d'établir les stratégies les plus appropriées pour dévitaliser avec une certitude presque totale l'ensemble des infrastructures de ses rivaux par l'entremise d'une attaque cybernétique ciblée à l'aide de codes malsains distribués d'une manière globale. Une telle attaque conduirait à une rupture des systèmes vitaux des États, à une mise hors service des réseaux électriques et de communication. En d'autres termes, les capacités de cyberguerre d'un État assisté par une superintelligence poseraient des enjeux de sécurité d'une ampleur inégalée en ce qui concerne le système international. En dehors même de la perspective d'une attaque physique lancée par un État assisté par une superintelligence, le risque d'une désinformation globale – ou sélective – des réseaux de communication serait une menace tout aussi préoccupante pour l'équilibre des relations diplomatiques et militaires. Il est clair qu'aujourd'hui une telle superintelligence n'a pas vu le jour. Néanmoins, nous savons que des intelligences artificielles spécialisées et entraînées sont présentement en mesure d'altérer le contenu des communications. De tels systèmes d'IA ne sont certes pas encore déployés mais leur développement est assurément en cours.

Une difficulté supplémentaire associée à la perspective – quoique lointaine – d'une superintelligence réside dans le fait qu'elle ne sera pas nécessairement qu'une affaire d'État. Une superintelligence, si elle voit le jour, émanerait très certainement d'une entreprise de haute technologie parvenue à maîtriser les briques constitutives d'une telle puissance de calcul. La compétition que viendraient à se livrer de tels acteurs suscite par ailleurs des préoccupations légitimes en matière de sécurité des systèmes d'IA qui seraient développés. La crainte des entreprises de se voir dépassées par leurs concurrentes dans cette course à la puissance technologique et aux champs applicatifs sur lesquels elle déboucherait inciterait ces entreprises à négliger les impératifs de sécurité et de stabilité de leurs produits. Le scénario le plus pessimiste serait celui d'une mise sur le marché de technologies d'IA instables, insuffisamment sécurisées et, peut-être imprévisibles. Quelle serait, dans une telle hypothèse, la validité des inférences produites par de tels systèmes ? Comment serait-on en mesure d'en évaluer la qualité ou la validité ? Cette éventualité n'est pas à prendre avec légèreté puisqu'elle s'appuie sur les leçons tirées des expériences passées de certains métasystèmes techniques. Il suffit, par exemple, de rappeler la substance des controverses émises dans le cadre du rapport sur l'explosion de la navette *Challenger* en 1986 pour s'apercevoir que les considérations en matière de prise de risque et de management n'aboutissaient pas obligatoirement aux décisions les plus « prudentes ».

Enfin, on ne saurait sous-estimer la possibilité pour un État/acteur voyou de parvenir à la mise au point d'un système d'IA performant et malveillant[65]. En effet, les ruptures technologiques ne sont pas uniquement le produit des activités de recherche menées dans des cadres académiques, gouvernementaux ou industriels. Compte tenu de la nature spécifique des technologies constituantes de l'intelligence artificielle, la possibilité pour un individu ou un groupe d'individus de parvenir à la conception d'une intelligence artificielle n'est pas nulle, même si elle demeure faible. En effet,

[65] Federico PISTONO, Roman V. YAMPOLSKIY, « Unethical Research: How to Create a Malevolent Artificial Intelligence », *cf.* https://arxiv.org/ftp/arxiv/papers/1605/1605.02817.pdf

le développement d'une IA « significative » suppose la maîtrise dans un large champ de compétences : compétences matérielles (obtention des composants hardware), logicielles (écriture d'algorithmes complexes) et acquisition de données. Cela suffit-il pour autant à ne pas devoir craindre la survenance d'un tel risque ? Aujourd'hui, les cyberattaques émanant d'individus isolés ou de groupes d'individus agissant de manière coordonnées sont légions. Et il n'est pas interdit de penser que l'IA représentera pour ces acteurs stratégiques d'un genre nouveau l'arme d'une de leurs offensives futures.

La question qu'il s'agit de poser est de déterminer si la perspective de l'avènement d'une superintelligence précipiterait la survenance d'une guerre préemptive contre le ou les acteurs parvenus au « seuil » d'une telle rupture technologique. Malgré les propos que peuvent tenir certains auteurs sur cette question, nous privilégions une approche plus prudente. Une certaine distance se doit d'être prise avec les scénarios les plus pessimistes développés à ce propos. Pourquoi ? Nombre d'hypothèses évoquant la possibilité d'une guerre préemptive conduite par des États suspectant l'un ou plusieurs de leurs homologues de parvenir (ou d'être parvenus) au seuil de la rupture technologique de la superintelligence sous-estiment les spécificités même de l'intelligence artificielle. Compte tenu des particularités d'une technologie telle que l'IA, il est peu probable que les systèmes de surveillance d'un État soient en mesure de détecter l'émergence d'une superintelligence, dans la mesure où celle-ci opérerait avec discrétion et n'émettrait aucun signal de son existence. Sans pour autant négliger la dimension matérielle de la compétition que se livrent les acteurs internationaux – publics et privés – en vue de maîtriser l'IA (sans même évoquer l'idée du superintelligence), il est fort probable que toute posture de guerre préemptive se fonderait principalement sur des considérations d'ordre psychologique, détachées donc de toute possibilité d'évaluer la dimension « matérielle » de la menace que poserait une IAG. Du reste, l'expérience historique regorge d'exemples de rapports de force qui ont rarement débouché sur des guerres préemptives déclenchées par des acteurs craignant d'être dépassés par l'armement de leur adversaire. En outre, là où des guerres préemptives furent conduites, ce fut souvent indépendamment de considérations matérielles prouvant l'existence d'un programme d'armement susceptibles de renverser l'équilibre des forces, quand il ne s'agissait pas simplement pour une puissance technologique dominante d'annihiler une organisation militaire qualitativement inférieure. Les mécanismes psychologiques incitateurs ou inhibiteurs d'une posture de guerre préemptive pourraient, du reste, être manipulés par des processus ayant recours à l'IA (sans qu'il s'agisse d'ailleurs ici d'IAG). De la même manière, des acteurs – qu'ils soient étatiques ou privés – déterminés à concevoir une IA à des fins malveillantes pourraient entreprendre de manipuler l'opinion publique et celle des décideurs politiques en engageant une campagne de désinformation globale et subversive via l'IA. Dans le domaine du contrôle et de la réduction des armements, nous avons assisté par le passé à la constitution de groupes de surveillance dont la mise en place répondait à la nécessité de prévenir et d'empêcher la conception et le déploiement de certains types d'armements. Ce fut par exemple le cas dans le domaine des armes chimiques, biologiques et bactériologiques mais également dans le champ des mines anti-personnel. La préoccupation première d'un acteur ou groupe d'acteurs animés par la volonté de constituer une IA à des fins malveillantes supposerait d'empêcher qu'un tel groupe de surveillance ne puisse jamais voir le jour. Une première stratégie consisterait à diluer des données contradictoires afin d'instiller le doute au sein de l'opinion à propos de la recherche dans le domaine de l'IAG. Ceci afin d'empêcher que des restrictions ne soient envisagées sur certains codes sources susceptibles de constituer des tremplins pour la recherche en matière d'IAG. Nous savons aujourd'hui à quel point certaines disciplines comme les sciences de l'évolution ou du climat sont attaquées de toutes parts au travers d'informations manipulées et détournées, voire d'infox, qui présentent – en apparence seulement – les dehors de la crédibilité scientifique. En s'employant à ce que l'opinion publique soit en proie au doute ou aux tergiversations, la manipulation des informations de base à propos de la recherche sur l'IAG permettrait d'empêcher l'élaboration de toute forme de moratoire en la matière.

Une seconde stratégie consisterait, pour cet acteur ou ce groupe d'acteurs résolus à avancer dans le champ de l'IAG, à convaincre ou menacer un ou quelques États. En assurant cet État ou ces quelques États de risques hors de proportion par rapport à la valeur qu'il(s) attache(nt) à l'élaboration d'un moratoire ou, à l'inverse, en promettant à ces États d'être les premiers à disposer de la puissance technologique fournie par l'IAG lorsque celle-ci verrait le jour, la possibilité de poursuivre les recherches en la matière serait garantie sur le plan international.

On le voit, l'accession au statut de « puissance maîtrisant l'IA » par des États désireux d'atteindre le niveau d'avancement technologique nécessaire et suffisant pour concevoir une IAG, même à des fins malveillantes, ne pourrait être que difficilement anticipée par leurs homologues. Cette difficile prévision de l'atteinte du seuil technologique diminue donc le risque de guerre préemptive mais réduit aussi, paradoxalement, les chances qu'un cadre normatif voie le jour.

2. La « guerre algorithmique »

Comme on peut le voir, les applications actuelles et futures de l'intelligence artificielle dans le champ militaire sont nombreuses et diverses. Certaines portent sur des technologies disruptives tandis que d'autres désignent davantage des capacités (à l'exemple de l'autonomie). Mais vers quel modèle de guerre les technologies de l'intelligence artificielle sont-elles sur le point de nous conduire ? Quelques observateurs des affaires militaires évoquent, pour qualifier l'interpénétration de l'IA dans le domaine militaire, l'émergence d'un nouveau type de guerre. Leur qualification peut différer mais la teneur des concepts présente de nombreuses similitudes. « Guerre algorithmique » et « hyper-guerre » sont les produits des principales réflexions menées sur le sujet. Pour rappel, un algorithme est une séquence d'instructions et de règles qu'emploient les machines dites « intelligentes » pour la résolution de problèmes spécifiques. Les algorithmes se situent au cœur du fonctionnement des systèmes d'IA et permet à ces derniers de transformer une masse d'informations en matière exploitable pour l'objectif visé. Peter Layton a conceptualisé cette expression issue du projet MAVEN porté par le Department of Defense (cf. plus loin).

L'émergence et la progression géométrique des performances des machines dites « intelligentes », fondées sur des processus d'apprentissage profonds via réseaux neuronaux, s'imposent aujourd'hui dans le calendrier d'acquisition des moyens dont se doteront les armées les plus avancées. Toutefois, en l'état actuel du développement technologique, il est utile de rappeler que les machines intelligentes dont il est question sont des intelligences artificielles spécifiques (*Narrow AI*) conçues pour égaler ou surpasser l'intelligence humaine dans des tâches particulières et limitativement définies. En d'autres termes, de telles machines intelligentes ont pour tâche d'appuyer la décision humaine et non de la remplacer. Les forces armées – on ne le répétera jamais assez – se montrent traditionnellement réticentes aux changements brusques et rapides que tente de leur imposer, à certaines époques, l'environnement sociotechnique. Dans le cas de l'IA, même si l'on suppose une IA faible, la rupture technique qui est présentée aujourd'hui au monde militaire dans une perspective d'intégration et de transformation des méthodes de conduite de la guerre implique l'acceptation d'une incompréhension des processus de traitement algorithmique, ce qui n'est pas sans poser un certain nombre de difficultés pour des autorités militaires dont l'exigence légitime est de pouvoir comprendre et expliquer les mécanismes – qu'ils soient humains ou computationnels – qui ont pu amener à la prise d'une décision particulière.

L'irruption du modèle de « guerre algorithmique » découle indiscutablement de la multiplication des dispositifs de senseurs et de récoltes des données issues des théâtres d'opération, zones de surveillance et espaces de bataille et dont le traitement des éléments colligés ne peut plus être assuré par les seuls opérateurs humains. Le volume des données produites annuellement dans le monde s'exprime aujourd'hui en zettabits. Une telle quantité de données exige l'appui de systèmes experts

en vue de leur analyse et de leur transformation en information exploitable par les hommes. Plus prosaïquement, l'exemple des méthodes d'analyse des données récoltées par le champ étendu de capteurs de l'US Air Force est particulièrement illustratif des limites auxquelles se heurte le personnel dans l'analyse des éléments issus des théâtres d'opération et zones d'intérêt stratégique pour les États-Unis. En l'état actuel des ressources humaines dont l'USAF dispose, il ne faut pas moins d'une vingtaine d'analystes œuvrant 24 heures sur 24 pour exploiter à peine 10 % de l'ensemble des données collectées par les multiples dispositifs d'acquisition. Les 90 % restant sont automatiquement stockés en l'attente d'un examen ultérieur… lorsque celui-ci est requis. Ce pourcentage de données traitées peut s'avérer encore moindre dans le cas de situations critiques ou de contraintes opérationnelles exigeant une analyse rapide de données restreintes. Cet exemple suffit à percevoir les limites d'un examen humain exhaustif des données essentielles pour la conduite des opérations militaires modernes.

3. Vers l'hyper-guerre ?

Les systèmes d'intelligence artificielle dont se dotent les armées des nations les plus technologiquement avancées mèneront, selon John R. Allen et Amir Husain, à l'avènement d'hyper-guerres. Dans un article désormais célèbre de la revue *U.S. Naval Institute Proceedings* paru en juillet 2017, les deux auteurs détaillent les contours de l'hyper-guerre et expliquent en quoi elle consacrera une modification substantielle non pas seulement des modalités d'exercice de la force, mais bien de la nature de la guerre elle-même. Toujours selon ces mêmes auteurs, l'idée communément admise selon laquelle la guerre, en accord avec la vision clausewitzienne, serait une « dialectique des volontés » est bientôt dépassée. Allen et Husain exposent les caractéristiques de l'hyper-guerre.

La première réside dans la mise en œuvre d'une capacité de commandement et de contrôle « distribuée » et « infinie ». Cette expression quelque peu sibylline signifie, plus concrètement, que le commandement et le contrôle ne dépendront plus de la cognition humaine, qui peut être sujette aux limites physiologiques des individus en charge de sa conduite. Désormais, les dispositifs de senseurs dispersés et réticulés en temps réel, combinés à des systèmes de traitement et d'analyse des données dotés d'une puissance de calcul inédite et en progression géométrique permanente, entraîneront une telle contraction de la boucle Observation-Orientation-Décision-Action que le tempo des opérations appuyées par l'intelligence artificielle exclura l'homme de la chaîne décisionnelle opérationnelle.

La seconde caractéristique de l'hyper-guerre est qu'elle assurera, selon les dires d'Allen et Husain, une coordination totale d'action et une parfaite conjonction des moyens. L'application d'une force supérieure à celle d'un adversaire ne suffit pas à remporter la victoire. Encore faut-il qu'elle soit dirigée de manière efficace sur un point précis et en temps utile. Une force armée en nombre inférieur peut ainsi faire la différence dans le cadre des hostilités pour autant que son commandement puisse être en mesure de produire les effets voulus en dirigeant les moyens sur le point le plus pertinent du dispositif adverse. Dans le cadre de l'hyper-guerre, la mise en réseau d'un grand éventail de senseurs permettra aux effecteurs de produire en un temps extrêmement réduit les effets désirés contre l'objectif identifié, et ce de telle sorte que ces effets entraîneront la décision. Une telle capacité de ciblage et de mise en œuvre des forces ne sera possible qu'à travers le recours à l'intelligence artificielle et exclura de sa boucle l'intervention humaine, dont les aptitudes cognitives se révéleront insuffisantes pour l'analyse en temps réel des forces en jeu et la définition appropriée des moyens d'action coordonnés.

Une troisième caractéristique de l'hyper-guerre est la simplification logistique. Les opérations militaires actuelles, en dépit de la place croissante des ordinateurs et réseaux dans l'organisation de la logistique et du support, comportent d'importantes latences que les moyens droniques et de recueil des données ne suffisent pas toujours à combler. L'opérateur humain exerce toujours la conduite et la

supervision des systèmes déployés et sa physiologie, déjà évoquée plus haut, peut affecter l'optimisation du traitement des informations.

De véritables véhicules autonomes embarquant des intelligences synthétiques intégrées au sein d'un vaste réseau, selon Allen et Husain, réduiront considérablement les « temps morts » opérationnels. Surtout, la combinaison de senseurs et d'effecteurs qui intégrera la plupart des systèmes de drones du futur assurera aux dispositifs de surveillance et de frappe une redondance sans précédent et garantiront une coordination logistique parfaite entre la détection de la cible, son analyse et son « traitement » par recours à des armes cinétiques ou à énergie dirigée.

Corollaire de l'intégration logistique et du commandement s'appuyant sur l'intelligence artificielle : l'adaptation en temps réel de la mission constitue la quatrième et dernière caractéristique de l'hyper-guerre. C'est là l'un des atouts principaux des systèmes d'armes équipés d'IA : la polyvalence parfaite d'une plate-forme dont la mission pourra instantanément changer selon les besoins opérationnels. Une IA apprenante présente un intérêt majeur : l'expertise acquise par un système donné peut faire l'objet soit d'un transfert instantané sur un système similaire, soit d'une modification immédiate du type de mission qui lui est confiée. Le processus d'apprentissage d'une IA ne souffre, en effet, d'aucune forme d'oubli qui serait liée, comme chez l'homme (même expert dans un domaine), au manque ou à l'absence de pratique. La connaissance expérimentale acquise par l'IA ne peut que tendre vers un renforcement perpétuel. C'est ce type d'apprentissage qui a permis à Alpha Go de tenir en échec des joueurs humains pourtant considérés comme des « experts » dans leur pratique du jeu de go.

Concluons notre exploration de l'hyper-guerre en insistant sur un point essentiel du développement exposé par Allen et Husain. Le modèle de l'hyper-guerre suppose une intégration étroite entre cybernétique, armes à énergie dirigée et force cinétique. Les frontières entre le virtuel et le réel s'effacent pour déterminer le mode le plus approprié d'effecteurs à mettre en œuvre compte tenu des données colligées et traitées par les systèmes d'IA interconnectés.

Au terme de cet exposé, une question s'impose : l'hyper-guerre constitue-t-elle une révolution dans les affaires militaires ? Les auteurs, Allen et Husain, prennent grand soin d'éviter le piège de la conceptualisation à outrance dans laquelle avaient pu verser certains de leurs prédécesseurs qui, en leur temps, avaient tenté de définir les contours de la guerre du futur résultant de l'explosion des technologies de l'informations, des ordinateurs et réseaux. La période qui suivit la guerre du Golfe de 1991 avait accouché d'une immense variété de concepts dont il serait ardu d'établir aujourd'hui un catalogue exhaustif.

Pour Allen et Husain, la nature même des technologies de l'IA n'affectera pas seulement les méthodes de conduite de la guerre moderne, elle altérera l'essence même de l'action humaine dans le champ guerrier. En d'autres termes, au-delà d'une révolution dans les affaires militaires, l'IA et les systèmes technologiques qui lui sont rattachés entraîneront une révolution dans les affaires humaines.

L'accélération géométrique du tempo des opérations générera un « effondrement » (*collapsing*) des fenêtres temporelles à l'intérieur desquelles chaque décision sera prise. Dans une telle configuration, l'assurance de la victoire ira, plus que jamais, au camp qui parviendra à recourir au plus grand nombre de systèmes de décision autonome et à engager le plus rapidement possible les moyens appropriés contre l'adversaire. Sur le plan stratégique, le commandement qui parviendra à s'appuyer sur des technologies de restitution de l'image du champ de bataille articulée autour de l'IA sera en mesure d'avoir une vision globale et en temps réel de l'évolution des opérations, ce qui lui garantira une capacité de prise de décision optimale, elle-même assistée par l'IA. Des assistants synthétiques intelligents ainsi que des dispositifs de réalité augmentée aideront le commandement à avoir une prise immédiate sur le cours des événements. Compte tenu de la nature et de l'ampleur des changements qui affecteront la conduite future des opérations militaires, il n'est pas certain que la guerre demeure

encore longtemps « une dialectique des volontés » au sens d'une opposition entre intelligences humaines. Les limites physiologiques de l'homme empêcheront sans doute ce dernier d'être en mesure d'appréhender la multiplicité des variables entrant en considération dans la prise de décisions qui relèveront de moins en moins d'instances humaines.

4. Les « flash wars » : le pire des mondes ?

Version paroxystique de l'hyper-guerre, le scénario d'une « guerre éclair cybernétique » – *flash war* – a fait son entrée dans la liste des risques auxquels pourraient aboutir notre dépendance aux systèmes d'IA pour la planification et la conduite des opérations militaires. Le modèle de la *flash war*, inspiré des *flash crashes* dans le domaine de la finance[66], associe à l'hyper-guerre l'incapacité de l'homme à s'immiscer dans la vitesse de calcul des processus qui viendraient à être engagés par des systèmes d'IA. Selon certains experts, le fait de confier à des systèmes d'IA des pans toujours plus étendus du processus décisionnel stratégique pourrait conduire à générer des rapports de forces sans la moindre possibilité d'établir des pauses dans l'enchaînement des décisions. En d'autres termes, le recours à des systèmes d'IA pour la préparation et la conduite de décisions stratégiques déboucherait sur une réaction en chaîne échappant totalement au décideur humain. L'univers physique de la guerre implique en général de nombreuses latences, que ce soit dans le cadre de la planification ou du déploiement. Les mouvements de troupes ou les préparatifs de systèmes de combat constituent des signaux que les protagonistes d'un conflit analysent afin de dégager un certain nombre d'indices concernant les intentions de l'adversaire. Dans un scénario de type *flash war*, ce sont principalement les systèmes cybernétiques qui constitueraient tout à la fois les opérateurs et les cibles. Il est très peu probable que la vitesse avec laquelle les attaques seraient conduites puisse permettre qu'interviennent des latences salutaires dans les processus décisionnels. Une *flash war* ne mettrait en œuvre que des systèmes cybernétiques, les systèmes d'armes physiques n'ayant pas le temps d'être activés.[67]

5. Le principe de dissuasion fait-il encore sens à l'ère de l'IA ?

Au travers des scénarios qui viennent d'être évoqués apparaît en filigrane la question de la pérennité du principe de dissuasion dans l'hypothèse où les principales puissances militaires des nations les plus avancées sur les plans scientifiques et technologiques viendraient à se doter de systèmes d'IA pour la planification militaire et la gestion des forces. Le thème de la dissuasion comporte deux volets : conventionnel et nucléaire. Nous nous attarderons ici sur la dimension nucléaire de la dissuasion tout en observant que celle-ci entretient des rapports constants avec les forces conventionnelles. La dissuasion nucléaire ne peut être considérée ex nihilo.

---

[66] Le concept de *flash crash* est né du krach boursier survenu en date du 6 mai 2010 aux États-Unis. Au cœur même de la crise de la dette grecque, l'indice boursier Dow Jones Industrial Average avait perdu près de 1000 points en près de 36 minutes. Ce phénomène était la résultante des systèmes de *high frequency trading* qui représentait alors près de deux tiers de l'ensemble des transactions opérées. Après cinq mois d'enquête, la Securities and Exchange Commission (SEC) et la Commodity Futures Trading Commission (CFTC) déposèrent un rapport conjoint le 30 septembre 2010 (intitulé *Findings Regarding the Market Events of May 6, 2010*), retraçant la séquence des événements ayant mené au *flash crash*. Le rapport présentait ainsi le « portrait d'un marché si fragmenté et fragile qu'une seule grande transaction pouvait faire partir les actions en spirale ». Il expliquait encore en détail comment une grande firme de fonds mutuel, vendant une quantité inhabituellement importante de contrats E-Mini (en) S&P 500, avait dans un premier temps épuisé les acheteurs disponibles, et comment ensuite les machines à algorithmes effectuant les transactions à haute fréquence (HFT) avaient vendu de manière agressive, ce qui contribua à accélérer l'effet de vente du fonds mutuel et avait contribué à la forte baisse de la valeur.

[67] Ulrike Esther FRANKE, *Flash Wars: Where Could an Autonomous Weapons Revolution Lead Us ?*, https://www.ecfr.eu/article/Flash_Wars_Where_could_an_autonomous_weapons_revolution_lead_us.

Quelles conséquences pourrait donc avoir le développement de l'IA en matière de dissuasion nucléaire ? Avant d'aborder les scénarios qui ont été ici et là élaborés par différents think tanks et groupes de travail sur ce sujet, il est utile de rappeler que, sans toutefois traiter expressément de la perspective d'une IA militaire, diverses réflexions ont par le passé vu le jour sur la question des rapports entre *dissuasion* et ce que l'on désignait alors par les technologies numériques. Dans un article de la revue *Defense Analysis* de 2001, Stephen Blank, professeur à l'U.S. Army War College, titrait cette interrogation de la manière suivante : « Can Information Warfare Be Deterred? »[68]. L'objectif de l'article de Stephen Blank était de scruter les similitudes possibles entre la dissuasion nucléaire et la dissuasion « informationnelle ». Revenant sur la doctrine de l'époque du Department of Defense américain laissant entendre qu'une supériorité informationnelle pourrait s'avérer tout aussi déterminante que la supériorité nucléaire qui caractérisait les forces armées des États-Unis, l'auteur de l'article tentait de mettre en lumières les perspectives et limites du principe d'*infodominance* avancé par les autorités politico-militaires américaines. Sans entrer dans les détails des propos de Stephen Blank, il convient de retenir les arguments présentés par ce dernier dans sa conclusion au sujet de l'impact de la dépendance de l'organisation militaire américaine aux ordinateurs et réseaux et d'en extraire des leçons susceptibles d'être étendues aux systèmes d'IA. Pour Stephen Blank, la dissuasion à l'ère de la guerre informationnelle s'avère particulièrement complexe et génératrice de risques nouveaux selon le type de rapport de forces considéré. L'une des principales difficultés auxquelles est confrontée la guerre informationnelle s'appuyant sur les technologies numériques est la validité de l'information circulant entre les protagonistes du rapport de force militaire. Un rapport de dissuasion entre protagonistes d'un rapport de forces militaires ne s'appuyant que sur la guerre informationnelle supposerait au préalable que l'ensemble des acteurs stratégiques fassent reposer universellement leur dissuasion sur les seuls moyens de guerre informationnelle. Il s'agit là, de toute évidence, d'une hypothèse d'école dont la probabilité de réalisation est extrêmement faible. Aussi, convient-il de nous interroger sur les paramètres de dissuasion d'un monde dans lequel coexisteraient (ce qui est effectivement le cas) des puissances disposant, pour certaines, d'une supériorité dans l'ensemble du spectre des moyens (informationnel, conventionnels, nucléaires, ADM), pour d'autres dans le seul domaine conventionnel, pour certaines encore dans le seul registre des ADM. Qu'en serait-il des principes régissant la dissuasion dans un tel système international ?

Compte tenu de l'incapacité pour une puissance militaire disposant de la supériorité informationnelle d'être en mesure de posséder une information totalement fiable à propos de l'attaque dont elle pourrait faire l'objet, une telle puissance militaire serait dans l'impossibilité d'identifier avec certitude son agresseur. Plus encore, pour les États dont les systèmes de force ne disposeraient pas d'une telle supériorité informationnelle, la tentation serait grande pour eux de recourir à la panoplie de leurs moyens ADM et asymétriques en guise de première frappe destinée à éviter toute forme de rupture de leurs systèmes d'information. Au lendemain de l'opération *Desert Storm* lancée à l'initiative des États-Unis pour repousser du Koweït les forces irakiennes de Saddam Hussein en 1991, le général indien Sundarji affirmait que quiconque prévoyait à l'avenir d'attaquer les États-Unis se devait de disposer d'un arsenal nucléaire. Autrement dit, dans un contexte stratégique où quelques acteurs disposent d'une supériorité informationnelle, la tentation sera grande pour certains acteurs de contrecarrer leur faiblesse en matière de systèmes informationnels par des attaques préemptives employant des moyens asymétriques, nucléaires ou ADM.

La question aujourd'hui posée est de déterminer si le développement de l'IA et son intégration au sein des systèmes de défense serait susceptible de précipiter le déclenchement d'un conflit et, plus spécifiquement, un conflit nucléaire. Cette hypothèse de travail a récemment été au cœur d'un rapport

[68] Stephen BLANK, "Can Information Warfare Be Deterred?", *Defense Analysis*, Vol. 17, No. 2, 2001, pp. 121-138.

produit par la RAND Corporation, qui exposait les résultats des échanges menés entre experts issus du monde militaire, de l'IA et des relations internationales à propos de divers scénarios. Pour comprendre le sérieux de l'hypothèse formulée par ce rapport, il convient de rappeler que, dans le cadre des derniers développements de la guerre froide en matière de commandement et de contrôle des forces stratégiques nucléaires (tant du côté soviétique que dans le camp américain), des systèmes de contrôle automatisés des armes nucléaires avaient été déployés dans l'éventualité où, en cas de frappes nucléaires, les éléments de commandement venaient à être entièrement détruits. Ainsi, l'Union soviétique avait-elle conçu le système *Perimetr*[69] (qualifié en russe de « Mertvaya Ruka », pouvant être traduit en anglais par « Dead Hand » ou en français par l'expression « Main Morte »). *Perimetr*, mis en service en 1985, était une illustration de dissuasion destructive. Le système pouvait enclencher automatiquement des missiles balistiques intercontinentaux par l'envoi d'un ordre préenregistré émanant de l'état-major des forces armées et du Commandement de gestion stratégique des Forces de fusées stratégiques à l'attention des postes de commandement et des silos individuels dans le cas où une attaque nucléaire venait à être détectée par des senseurs sismiques, de rayonnement ou de radioactivité et de surpression. En d'autres termes, *Perimetr* devait assurer la continuité de la stratégie nucléaire de l'Union soviétique par le déclenchement automatisé de représailles même en cas de destruction de toute capacité de décision humaine pour le lancement des ICBM[70]. Le principe de fonctionnement de *Perimetr* reposait sur une structure à deux composantes : le missile de commandement et le système de commande et de contrôle autonome. Dans l'hypothèse d'une frappe nucléaire contre le territoire et les forces armées de l'URSS, un missile de commandement (une fusée 15P011 disposant d'une ogive radio 15B99) avait pour mission d'assurer la transmission de l'ordre général de lancement à l'ensemble des postes de commandement et complexes de lancement des missiles. Le cœur de *Perimetr* était, toutefois, le système de commandement et de contrôle autonome. Il est supposé qu'il disposait d'une structure équipée d'instruments capables de suivre la présence et l'intensité des communications sur fréquence militaire, de recevoir les signaux télémétriques des postes de commandement, de mesurer le niveau de rayonnement à la surface et de déterminer les sources de ce rayonnement à proximité afin de les comparer aux mesures provenant des détecteurs de perturbations sismiques. Après avoir procédé à la comparaison de l'ensemble de ces données, le système de corrélation était la dernière étape avant le lancement des missiles nucléaires. La raison pour laquelle *Perimetr* reçut l'appellation de « Dead Hand » tenait à la probabilité de l'existence d'un interrupteur « homme mort » sur le système. Cet interrupteur pouvait être activé par le commandement suprême de telle sorte que si le système ne détectait pas de signal d'arrêt de la séquence algorithmique de combat, le lancement automatique des missiles était ordonné. L'objectif d'un système tel que *Perimetr* était de convaincre les États-Unis de la fermeté – et de l'inéluctabilité – de la réponse nucléaire soviétique dans l'hypothèse d'une crise majeure entre les deux superpuissances. Face au projet américain d'Initiative de défense stratégique (IDS) et du programme de bouclier antimissiles qu'il comportait, l'Union soviétique avait choisi l'option de l'ultra-déterminisme pour le type de riposte à produire en cas de frappes nucléaires américaines, frappes dont elle jugeait la probabilité de survenance très élevée depuis le retour de l'administration républicaine à la Maison Blanche. *Perimetr* avait donc pour mission d'assurer la frappe de riposte en toutes circonstances.

Du côté américain, un programme similaire avait également vu le jour : l'AN/DRC-8 Emergency Rocket Communication system (ERCS). L'objectif de l'ERCS, très similaire à *Perimetr*, était de garantir un système de communication stratégique pour la United States National Command Authority.

[69] Littéralement, en russe, le dispositif était appelé « Système de périmètre ».

[70] Edward GEIST, Andrew J. LOHN, (eds), *How Might Artificial Intelligence Affect the Risk of Nuclear War?*, RAND Corporation, Santa Monica (Calif.), Serie Perspective, 2018.

Le système reposait sur le lancement d'une ogive radio UHF installé, en lieu et place d'une ogive nucléaire, sur un missile Minuteman II. Dans l'éventualité d'une frappe nucléaire, le système procédait au lancement exo-atmosphérique (en couche basse) du transmetteur pour la diffusion d'un message d'alerte à l'attention des unités du Strategic Air Command.

6. Shall We Play A Game ?

Tant du côté soviétique que dans le camp américain, l'idée de procéder à une délégation complète de la décision de frappe à une machine a rencontré de nombreuses réticences, surtout en raison du fait que les commandants humains – qu'ils soient politiques ou militaires – ont toujours souhaité se réserver une telle habilitation. Le système *Perimetr* est la preuve du poids de telles réserves envers une automation du déclenchement du feu nucléaire, même en mode riposte. Son surnom de « Dead Hand » témoigne surtout du fait qu'un abandon total de la décision de lancement des missiles ICBM n'a jamais rencontré de consensus au sein des états-majors politico-militaires. La décision ultime se situait toujours entre les mains d'un opérateur humain. Et même en cas d'activation éventuelle de l'interrupteur « Dead Hand », les sources convergent pour rapporter que le choix de cette activation demeurait le fait d'un commandant humain : la procédure « Dead Hand » ne faisait que garantir la pérennité de la décision de lancement dans l'hypothèse d'une apocalypse nucléaire sur le sol soviétique. Par ailleurs, les exemples d'erreur d'évaluation des systèmes d'automation en matière d'évaluation de la menace sont légions dans l'historiographie de la guerre froide. En novembre 1979, un exercice de guerre chargé par erreur sur un ordinateur du Commandement de la défense aérospatiale de l'Amérique du Nord (North American Air Defense Command – NORAD) avait fourni au système des informations – évidemment erronées – à propos d'une attaque nucléaire soviétique imminente. C'est la vérification par le NORAD de son système radar qui permit de déceler l'erreur et d'avertir les opérateurs de la fausse alerte[71]. Cette procédure de double vérification dans le cas d'une alerte nucléaire, appelée « dual phenomenology », avait précisément pour but de conduire deux analyses indépendantes d'un même signal indiquant l'imminence d'une attaque nucléaire. Cette méthode, bien évidemment coûteuse, n'existait pas nécessairement du côté soviétique. En juin 1980, ce fut le conseiller à la sécurité nationale, Zbigniew Brzezinski, qui fut réveillé par un appel du NORAD à 2 h 26 du matin l'avertissant de l'attaque imminente de 220, puis de 2 200 missiles nucléaires soviétiques. Le conseiller Brzezinski s'apprêtait à réveiller le président Jimmy Carter de l'alerte pour lui intimer de lancer une riposte en quelques minutes lorsqu'un troisième appel du NORAD lui parvint afin de lui indiquer qu'il s'agissait d'une fausse alarme. L'erreur d'évaluation du système d'automation sembla provenir d'un processeur défectueux (d'une valeur d'à peine 1 dollar US) d'un ordinateur relié au système informatique du NORAD. En 1983 enfin, en Union soviétique, un système d'alerte avancée rapporta par erreur le lancement de cinq ICBM américains avant qu'un officier en poste, le lieutenant-colonel Stanislav Petrov, ne procède à une manœuvre de correction, suspectant une erreur de la part du système. Son appréciation s'avéra juste puisque ce que le système informatique avait considéré comme des indicateurs de départ de missiles étaient en fait le reflet de la lumière solaire dans des nuages[72].

[71] Cet événement inspira les réalisateurs du film *Wargames*, sorti en 1983, dont le personnage central, un jeune féru d'informatique, avait piraté involontairement le supercalculateur du NORAD et lancé à travers celui-ci une simulation de compte à rebours pour le lancement d'une attaque nucléaire contre l'Union soviétique. L'intitulé de cette section (Shall We Play A Game ?) est la question célèbre que pose le supercalculateur au héros du film.

[72] John BORRIE, « Cold War Lessons for Automation in Nuclear Weapon Systems », in Vincent BOULANIN, (eds), *The Impact of Artificial Intelligence on StrategicStability and Nuclear Risk*, *Volume I: Euro-Atlantic Perspectives*, Solna, Stockholm International Peace Research Institute (SIPRI), may 2019, pp. 41-52.

Les États-Unis semblent avoir exploré plus en avant la possibilité d'adjoindre une IA à la procédure de lancement de missiles nucléaires en cas de crise mettant en péril les intérêts vitaux du pays. À la fin des années 1980, le Survivable Adaptive Planning Experiment (SAPE) permit aux forces armées des États-Unis d'étudier l'emploi de technologies issues de l'IA pour permettre aux forces stratégiques de cibler les unités mobiles de lancement ICBM soviétiques. Le programme SAPE, néanmoins, ne disposait pas du contrôle direct sur les armements nucléaires ; il était plutôt conçu comme un système expert chargé de traduire les données de reconnaissance au sein de plans de ciblage des missiles nucléaires pour que ceux-ci soient ensuite transportés par des bombardiers B-2 pilotés.

L'histoire militaire de ces 70 dernières années le montre : le débat sur l'automation et la délégation aux machines des fonctions de prise de décision de frappes – y compris les frappes incluant des moyens nucléaires – n'a cessé de figurer sur l'agenda des états-majors des nations les plus avancées sur les plans scientifique et technologique et, tout particulièrement, parmi les puissances nucléaires. À l'origine, l'idée d'inclure des dispositifs automatisés ou semi-automatisés au sein des processus de décision de lancements de missiles répondait principalement à la nécessité de garantir une capacité de frappe en second. Il ne fut, semble-t-il d'après les sources ouvertes disponibles, jamais question de confier à des ordinateurs la moindre capacité de délibération d'une attaque en premier. Comme a pu le rappeler Benjamin Hautecouverture, l'Histoire délivre quatre éléments essentiels de réflexion sur ce sujet[73] :

1. Les progrès intervenus dans l'apprentissage machine et dans l'automation n'ont pas ouvert un champ inédit de la réflexion stratégique, pas plus qu'ils n'ont auguré l'imminence de nouvelles possibilités opérationnelles.
2. La principale préoccupation des états-majors des pays intéressés par la question de la place à accorder aux dispositifs d'automation résidait dans une meilleure connaissance des systèmes que développait l'adversaire en la matière. Le souci des élites politico-militaires était de comprendre dans quels buts et selon quelles procédures de semblables systèmes déployés par l'adversaire étaient orientés. L'exemple du système soviétique *Perimetr* est particulièrement éclairant. Tandis que, pour les forces armées soviétiques, *Perimetr* offrait un gage de stabilité dans la dissuasion nucléaire en évitant de confier à un dirigeant la prise de décision ultime en situation de crise, du côté occidental, l'opacité du système présentait l'image d'un dispositif tout aussi automatisé que déstabilisant.
3. Quelles que furent les périodes durant lesquelles de tels systèmes d'automation furent envisagés et parfois mis sur pied, ils démontrèrent régulièrement des insuffisances liées en grande partie à des problèmes de maturité des technologies. Il résulta donc à chaque fois de ce constat la formation d'un consensus autour de l'importance du contrôle humain et de la redondance des procédures de vérification des traitements de données.
4. D'une manière générale, les systèmes de force qui ont tenté d'incorporer au sein de leurs architectures des dispositifs d'automation ou d'intelligences artificielles se sont révélés particulièrement hésitants, voire réticents à cette perspective. Ce sont en général des critères de sûreté et de responsabilité qui ont prévalu. Il s'agit par ailleurs de ne pas ajouter aux systèmes de forces des vulnérabilités nouvelles et d'empêcher que des failles informatiques ne viennent fragiliser la fiabilité des architectures de forces.

[73] Benjamin HAUTECOUVERTURE, « Applications nucléaires de l'automatisation : rappels historiques », *Bulletin mensuel de l'Observatoire de la dissuasion*, sous la direction de Emmanuel MAÎTRE et Bruno TERTRAIS, Paris, Fondation pour la recherche stratégique (FRS) & Direction générale des relations internationales et de la stratégie (DGRIS), numéro 67, été 2019, p. 13.

L'interrogation qu'il nous reste à explorer au terme de ce rappel historique et des leçons tirées de la guerre froide est la suivante : les avancées récentes intervenues dans le domaine de l'IA sont-elles de nature à redéfinir le cadre stratégique de la dissuasion nucléaire ? On rappellera avant tout que, comme le confirme l'histoire des relations stratégiques entre les puissances nucléaires depuis la fin de la Seconde Guerre mondiale, l'enjeu de l'automation au sein des capacités de frappes nucléaires a constitué une thématique récurrente des débats stratégiques. Aujourd'hui, les performances nouvelles qu'offre l'IA – sans que les instances militaires ne soient parvenues pour autant à en définir le concept – ont remis au goût du jour un certain nombre de réflexions. Comme nous l'avons indiqué, plusieurs organismes de réflexion, et non des moindres, ont replacé la question de l'IA au cœur de leurs travaux. Le cas du Stockholm International Peace Research Institute (SIPRI) est particulièrement révélateur puisque deux récentes publications majeures ont été produites sur ce sujet, soit trente ans après l'ouvrage d'Allan M. Din, édité également par le SIPRI et Oxford University Press, et intitulé « Arms and Artificial Intelligence: Weapons and Arms Control Applications of Advanced Computing » (1987). La résurgence du thème de l'IA et de la possibilité d'intégrer celle-ci à l'architecture de commandement et de contrôle des armements nucléaires résulte donc des apparentes prouesses nouvelles des technologies sur lesquelles elle repose. Cependant, les inconnues qui entourent la dynamique du récent « sursaut » de l'IA ne sont toujours pas levées : ces progrès sont-ils graduels ou procèdent-ils de sauts discontinus ? Doit-on s'attendre encore à la survenance de cycles de développement et, par conséquent, à la possibilité de nouvelles stagnations des recherches et des réalisations (autrement dit, de nouveaux hivers de l'IA) ? Faut-il s'attendre à l'irruption d'une « superintelligence » génératrice de transformations socio-politiques, économiques, juridiques, éthiques et militaires sans commune mesure ? Quelles que soient les réponses que nous apportons à ces questions (qui sont autant de postures intellectuelles sur les caractéristiques et les perspectives de l'avènement de l'IA), la place que pourrait occuper demain l'intelligence artificielle au sein des systèmes de forces dépendra du cadre théorique, culturel et doctrinal dans lequel elle sera intégrée.

En ce qui concerne la dissuasion en tant que telle, la place de l'IA dans le champ nucléaire doit s'analyser à quatre niveaux. Un premier niveau est celui de l'alerte avancée et de la collecte du renseignement. En la matière, l'apprentissage machine est utilisé depuis de nombreuses années avec succès. Les systèmes qui équipent les organismes d'analyse politico-militaires américains établissent d'ores et déjà des corrélations entre paquets de données issues de dispositifs de récolte hétérogènes. Ceci est particulièrement vrai dans le domaine de la lutte contre le terrorisme. L'apprentissage machine est également employé pour l'identification rapide des déploiements de forces et la connaissance de situations évolutives sur des théâtres distants. Une mise à disposition plus rapide des données en provenance des senseurs déployés permet d'offrir un temps plus étendu à la décision humaine : une approche qui relativise quelque peu l'imaginaire de science-fiction à propos d'un remplacement de la décision humaine par la machine. Un second niveau d'analyse est l'apport de l'apprentissage machine et de l'IA dans le domaine des dispositifs de commandement et de contrôle. Bien qu'une part importante de la littérature spécialisée se concentre sur cet aspect de la décision militaire, il convient d'indiquer que peu de transformations significatives sont à prévoir dans ce secteur, et ce pour deux raisons. La première est que l'automation est une technique déjà développée dans ce domaine. La seconde est que les algorithmes développés par les ingénieurs ne témoignent pas d'une fiabilité suffisante au regard de la criticité de ce domaine. Il n'en demeure pas moins que l'intelligence artificielle offre des percées réelles en matière de cybersécurité pour la protection des systèmes. Un troisième domaine applicatif est le contrôle des vecteurs de frappe. Ici, des recherches exploratoires s'appuyant sur les dernières avancées de l'IA se font jour et pourraient déboucher sur un certain nombre de ruptures. Celles-ci concerneront sans doute la précision des engins et l'adaptation de leur manœuvrabilité (en ce qui concerne notamment les vecteurs hypersoniques) pour une meilleure assurance de la dissuasion. Enfin, un quatrième niveau d'analyse concerne le ciblage par

les systèmes conventionnels défensifs et en matière cybernétique (offensif et défensif) ou encore de guerre électronique. Au terme de cette exploration, il apparaît que les technologies d'IA pourront certes apporter une plus grande sécurité et sans doute une meilleure réduction du risque nucléaire. Il importe toutefois de relativiser la portée révolutionnaire de l'IA dans ce secteur en précisant que les améliorations progressives qu'elle autorisera ne déboucheront pas, à un horizon prévisible, à des ruptures technologiques réelles qui pourraient remettre en question les fondamentaux de la stratégie.

Ceci étant dit, des réserves doivent toutefois être formulées qui conduisent à indiquer que des changements profonds, encore difficilement décelables, pourraient survenir sur une période allant du moyen au long terme. Ces changements concerneront tout d'abord, même s'ils ne s'inscrivent pas entièrement dans le champ de la dissuasion nucléaire (tout en y étant associés), la problématique de la défense antimissile et de la reconnaissance stratégique. L'un des arguments avancés de manière récurrente par les détracteurs des systèmes de défense antimissiles réside dans l'impossibilité d'une prédictibilité parfaite des vecteurs balistiques engagés par un État agresseur. Il existe, en effet, dans le chef d'un État qui déciderait de lancer une attaque balistique contre une puissance dotée de systèmes antimissiles une panoplie de moyens susceptibles d'altérer les chances d'interception. On pense notamment à des frappes de saturation, des vecteurs hypersoniques ou hypermanœuvrants ou, tout simplement, le recours à des dispositifs de leurre de dernière génération. D'une manière générale, la principale difficulté pour un État disposant d'une défense antimissile balistique serait de s'assurer que sa contre-offensive garantisse une interception du vecteur tout en disposant d'une évaluation aussi crédible que possible des conséquences liées à cette interception en termes de dommages collatéraux liées aux retombées des débris du vecteur intercepté (plus encore lorsqu'il s'agit d'une frappe à l'aide de missiles équipés de charges CBRN). Un système d'intelligence artificielle pourrait contribuer à l'efficacité accrue d'une architecture de défense antimissile en permettant une meilleure reconnaissance de cible, une plus grande précision dans le calcul des trajectoires des ogives (notamment à l'encontre de missiles mirvés[74]) et l'appréciation des risques de retombées. Ce faisant, une défense antimissile appuyée par une IA de traitement conduirait à une altération considérable de la probabilité de succès d'une première frappe que viendrait à engager un État hostile doté de moyens balistiques. Ce déséquilibre serait de nature à générer une course aux armements et à encourager le recours à des frappes de préemption par un État soupçonnant son adversaire d'être sur le point de développer un système de défense antimissile appuyé par une IA[75].

Cette observation sur la défense antimissile nous amène à considérer l'impact de l'IA sur la notion de vulnérabilité qui constitue, comme on le sait, le cœur de la dissuasion nucléaire. Pour qu'un rapport de dissuasion entre deux puissances nucléaires puisse s'établir, il doit exister une vulnérabilité partagée de telle sorte que l'État effectuant la première frappe puisse à son tour faire l'objet d'une riposte destructive. Or, les avancées récentes dans le domaine de l'IA en matière de reconnaissance et d'imagerie pourraient remettre en question ce principe, notamment en offrant à l'un des protagonistes du rapport de dissuasion une capacité bien supérieure d'analyse des données issues des senseurs. Grâce au traitement par IA des informations issues des images satellites combinées aux données recueillies par les systèmes de reconnaissance déportés en profondeur sur le territoire

[74] Le mirvage désigne une technique permettant à un missile d'être équipé de plusieurs têtes (nucléaires ou conventionnelles) qui suivent chacune une trajectoire indépendante dès leur rentrée dans l'atmosphère. Cette technologie fut introduite pour la première fois en 1968 lors de tests opérés sur les missiles américains Minuteman III. Le déploiement de missiles mirvés, originellement destinés à contrecarrer les systèmes de défense antimissiles de l'Union soviétique, a considérablement modifié l'équilibre des forces entre les États-Unis et l'Union soviétique. La technologie du mirvage a conduit les deux superpuissances à ériger la destruction mutuelle assurée (MAD) comme le fondement de la dissuasion.

[75] Li XIANG, « Artificial Intelligence and Its Impact on Weaponization and Arms Control », in Lora SAALMAN (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Solna (Sweden), Stockholm International Peace Research Institute (SIPRI), octobre 2019, p. 13.

adverse (drones MALE[76], drones HALE[77]), une puissance nucléaire pourrait disposer d'une connaissance pointue et en temps réel des moyens de frappe de son adversaire et être tentée de procéder à une première frappe désarmante avec une certitude élevée de contraindre son adversaire à renoncer à toute riposte. Ce faisant, l'IA conjuguée aux capacités de frappes nucléaires aura pour effet d'affaiblir la stabilité stratégique entre deux puissances nucléaires unies par un rapport de dissuasion.

Toutefois, le scénario dans lequel l'impact de l'IA pourrait apparaître comme le plus déstructurant pour la dissuasion fait intervenir les capacités de frappe depuis des sous-marins lanceurs d'engins. La stabilité de la dissuasion nucléaire, nous l'avons dit, repose sur la préservation d'une vulnérabilité mutuelle entre les protagonistes de ce rapport de dissuasion. Les moyens de frappe depuis des sous-marins lanceurs d'engins assurent une option de riposte (seconde frappe) à une puissance nucléaire dont le territoire, les forces terrestres ou les villes viendraient à être l'objet d'une première frappe. La possibilité d'une riposte provenant d'une composante sous-marine bénéficiant de la discrétion que lui confère la nature même de son système d'armes génère pour la puissance qui viendrait à envisager une première frappe une crainte d'une ampleur telle qu'elle inciterait au renoncement de toute agression. Or les systèmes de traitement et d'analyse des données par IA dans le contexte de la lutte anti-sous-marine généreront à n'en pas douter une déstabilisation des équilibres stratégiques. Une IA spécifiquement entraînée à l'étude et à l'analyse de la multitude de données issues des senseurs océanographiques et des moyens de lutte anti-sous-marine annihilerait l'atout essentiel de l'arme sous-marine dans le cadre de la dissuasion : sa discrétion absolue. Ici encore, on voit poindre le problème pour l'équilibre d'un rapport de dissuasion. En mettant un État adverse dans l'impossibilité de garantir la furtivité de sa composante sous-marine, affectant ainsi sa capacité de riposte nucléaire par frappe sous-marine , une puissance nucléaire dotée d'un système d'IA serait tentée de procéder à une première frappe désarmante contre les moyens sous-marins de l'adversaire[78]. Cette hypothèse est très sérieusement prise en compte par la Chine dans son rapport de dissuasion avec les États-Unis. Pékin, conscient de l'infériorité relative de son arsenal nucléaire et de l'avance acquise par les États-Unis en matière d'intelligence artificielle dans le champ stratégique, a considérablement investi dans le déploiement d'unités mobiles de frappe balistique et de camouflage de ses capacités de frappe. Elle mise également sur des stratégies de manœuvrabilité et de survivabilité de ses systèmes. Néanmoins, si les États-Unis confirment leur recours à des dispositifs d'IA pour l'analyse des données, une incertitude réelle pèsera sur les capacités nucléaires mobiles chinoises et le risque d'une première frappe désarmante en provenance des États-Unis sera jugée comme hautement probable. On le voit : la maîtrise de l'IA par un seul des deux États liés par une relation de dissuasion suffit à conduire à un accroissement de l'instabilité de cette relation. Dans le cas que nous évoquons, il est certain que la Chine augmentera son niveau d'alerte et se placera dans une posture de frappe préemptive tandis que les États-Unis, convaincu de leur accession à une analyse parfaite et exhaustive des données des senseurs grâce à l'IA, seront tentés par le principe d'une première frappe désarmante.

Les États-Unis disposent-ils aujourd'hui concrètement d'une capacité d'analyse des données appuyée par l'IA ? Il semble que cette capacité soit effectivement maîtrisée par le Département de la Défense. En 2017, des chercheurs de l'Université du Missouri ont publié un article portant sur l'application de modèles d'apprentissage profond en matière de reconnaissance d'images satellites. Dans cet article, les chercheurs ont expliqué avoir entraîné une IA en employant pas moins de 2 200 images satellites de positions de systèmes de missiles sol-air. En à peine 42 minutes, l'IA a été en mesure d'identifier

[76] Medium Altitude Long Endurance.
[77] High Altitude Long Endurance.
[78] Edward GEIST, Andrew J. LOHN, (eds.), *op. cit.*, Santa Monica (Calif.), RAND Corporation, Serie Perspective, 2018, p. 10.

avec une précision évaluée à 90 % des positions de missiles antimissiles chinois ; un travail qui aurait pris plus de 60 heures à un analyste humain[79]. En dépit des difficultés qu'a rencontrées l'IA dans cette expérience pour identifier des positions camouflées, la prouesse réalisée par les chercheurs de l'Université du Missouri a conduit l'US Office of Naval Research (ONR) à rédiger le cahier des charges d'un futur programme appelé à faire traiter par une IA l'ensemble des relevés physiques océanographiques et sonores marins ainsi que leurs mises à jour afin de permettre à cette IA d'apprendre à identifier les perturbations sous-marines indiquant la présence de submersibles[80].

7. De l'équilibre entre alerte avancée et prise de décision

Le rôle et la place de l'IA au sein des dispositifs militaires en charge de la dissuasion nucléaire (de l'alerte avancée à la frappe opérée) ne dépendent pas uniquement des performances techniques des machines. Leur contribution résultera également de la manière dont les éléments humains intégrés au processus interpréteront les données traitées par l'IA. Ceci ne vaut pas uniquement dans l'hypothèse d'une erreur commise par l'IA, mais également dans l'éventualité d'un fonctionnement optimal de la machine dans le cadre des données qu'elle aurait à traiter. Il importe néanmoins de s'interroger sur la façon dont ces systèmes d'IA contribueront à formater, demain, la faculté de perception et d'interprétation des décideurs humains. Le substrat historico-culturel des sociétés n'est pas sans conséquence sur la vision des hommes à propos du rôle de l'IA, y compris dans le volet militaire.

Lora Saalman, de l'East-West Institute, explore cette question à propos du discours de la Chine en matière d'IA[81]. Dans un premier temps, Lora Saalman précise l'angle selon lequel Pékin aborde l'IA dans son discours. Contrairement aux visions occidentales, le gouvernement chinois envisage l'IA selon une approche bien plus pragmatique. Le discours chinois autour de l'IA ne comporte pas de référence aux notions occidentales de « *singularité* », de « superintelligence » qui opèrent surtout en Occident comme autant d'interférences dans un débat qui devrait s'affranchir de toute forme de sensationnalisme. Et les rares évocations par Pékin de ces notions occidentales ne sont là que pour caractériser la manière dont l'IA est perçue en dehors de la Chine. Lora Saalman revient, dans un second temps, sur les préoccupations qui se situent à la base de l'approche chinoise de l'IA. Il existe, selon elle, une différence fondamentale dans les approches qui guident les perceptions américaine et chinoise d'une attaque nucléaire. Les forces armées des États-Unis, selon l'auteure, craignent la survenance d'un « *false positive* », tandis que les Chinois redoutent la possibilité d'un « *false negative* ». Expliquons cette différence. Aux États-Unis – comme les différents incidents liés aux fausses alertes à la charnière des années 1970 et 1980 l'ont démontré – la principale crainte nourrie à l'encontre du processus de dissuasion réside dans le déclenchement d'une riposte nucléaire à la suite de la détection d'une attaque non avérée. À l'inverse, du côté chinois, c'est la crainte d'une faille dans les capacités techniques de détection des départs de missiles adverses qui mobilise les états-majors. Cette obsession de la frappe soudaine et non détectée (*bolt-from-the-blue attack*[82]) se situe à la base de la démarche chinoise visant le développement tous azimuts de l'IA pour des besoins militaires. On remarquera, par ailleurs, que l'IA ne constitue pas en soi un axe de Recherche & Développement au sein des différents services de l'appareil de défense chinois. Il s'agit plutôt pour la Chine d'irriguer les multiples constituants de son organisation militaire par le renfort d'applications à base d'IA et de

[79] Richard A. Marcum, Curt H. Davis, Grant J. Scott & Tyler W. NIRVIN, « Rapid Broad Area Search and detection of Chinese Surface-to-Air Missile Sites Using Deep Convolutional Neural Networks », *Journal of Applied remote Sensing*, vol. 11, no. 4, octobre-décembre 2017.

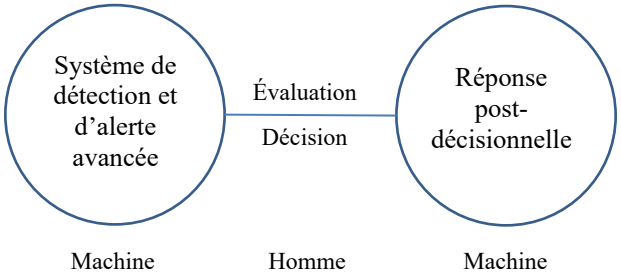
[80] Patrick Tucker, « How AI Will Transform Anti-Submarine Warfare », *Defense One*, 1er juillet 2019.

[81] Lora Saalman, « Fear of False Negatives: AI and China's Nuclear Posture », *Bulletin of Atomic Scientists*, 24 avril 2018, *cf*. https://thebulletin.org/2018/04/fear-of-false-negatives-ai-and-chinas-nuclear-posture/.

[82] Expression que nous pourrions traduire par « coup de tonnerre dans un ciel bleu ».

systèmes autonomes. L'obsession de Pékin demeure le risque de passer à côté de la détection d'une attaque et c'est à l'aune de ce « biais » de culture stratégique (lui-même le produit d'une histoire marquée par des revers découlant de manœuvres de surprise dont elle fut victime sur les plans politique et diplomatique) qu'il importe de comprendre le plan de développement d'une intelligence artificielle de nouvelle génération publié par le pays au mois de juillet 2018.

Cet exemple de différence de culture stratégique expliquant les formes particulières de l'implantation des systèmes d'IA au sein de leurs forces armées respectives nous renvoie au poids des modalités d'interprétation des données recueillies et traitées par l'IA au sein des organisations militaires. Pour comprendre les enjeux C3I (*Command, Control, Communications and Intelligence*) en matière de stratégie nucléaire, on pourrait représenter l'architecture de commandement comme une construction composée de deux ensembles en liaison. Le premier ensemble regrouperait les systèmes de détection et d'alerte avancée tandis que le second comporterait les mécanismes de réponse post-décision. Il importe de réaliser qu'au sein de chacun de ces ensembles de dispositifs figurent des processus pour l'essentiel automatisés. Si l'on considère, par exemple, la modalité de réponse post-décisionnelle que constitue le lancement d'un missile, il s'avérera compliqué – sinon impossible – d'en interrompre la séquence et le parcours. L'interrogation qu'il s'agit de poser consiste à se demander quelles pourraient être les conséquences d'une intégration de l'IA dans chacun des ensembles que sont, d'une part, la détection et l'alerte avancée et, d'autre part, la réponse post-décisionnelle ? En d'autres termes, le développement de l'IA au sein des dispositifs de détection et d'alerte avancée pourrait-il entraîner un débordement de cet ensemble sur les processus d'évaluation et de décision au point de conduire à un retrait total de l'homme de ce schéma d'action ? De nombreux experts qui pourtant consentent à inclure des systèmes d'IA au sein du processus de détection et d'alerte avancée insistent néanmoins sur la nécessité de préserver un sanctuaire à l'intervention humaine au sein de la chaîne de commandement et de contrôle : il est essentiel que les deux ensembles n'aient aucun contact entre eux. Ceci étant posé, on soulignera que la seule présence humaine au sein d'un processus décisionnel largement automatisé, voire autonomisé du fait de l'intégration d'une IA à travers les différents maillons de la chaîne de commandement ne saurait suffire à empêcher l'occurrence d'une erreur au niveau du traitement des données par les machines. L'homme peut lui-même constituer, en dépit de son humanité physiologique, le maillon d'un processus automatisé. L'expérience de Stanley Milgram (conduite entre 1960 et 1963) a pu attester du degré d'obéissance dont peut faire preuve un individu devant une autorité qu'il juge légitime. Il n'est pas interdit de penser que, dans une chaîne de commandement intégrant un système d'IA, un opérateur humain, convaincu de la légitimité de l'IA et des données analysées par elle, se retrouve dans une situation où il n'est pas en mesure de déroger aux injonctions délivrées par l'IA et paralyse toute réinterprétation humaine de la situation de crise.



8. IA et commandement : vers une fuite en avant ?

La récente proposition de budget pour l'année fiscale 2021 déposée par le Department of Defense américain laisse toutefois songeur quant aux perspectives d'évolution des systèmes de commandement et de contrôle dans le domaine de la dissuasion nucléaire. Au sein de cette proposition, une enveloppe de 28,9 milliards de dollars américains est consacrée à la modernisation du complexe d'armements nucléaires, soit le double du montant qui fut demandé pour 2020. Dans cette nouvelle enveloppe figurent certes l'acquisition de nouveaux vecteurs nucléaires dont le nouveau bombardier B-21 *Rider* (2,8 milliards de dollars), le sous-marin nucléaire lanceur d'engins de classe *Columbia* (4,4 milliards de dollars) et un nouveau missile balistique intercontinental (*Long-Range Stand-Off* – LRSO, pour un montant de 474 millions de dollars). Dans le cadre de la modernisation du dispositif de dissuasion nucléaire, on soulignera cependant l'évocation d'un montant de 7 milliards de dollars pour le Nuclear Command, Control and Communications (NC3). Il faut encore ajouter à cela les 841 millions de dollars proposés pour les systèmes d'intelligence artificielle[83]. Des sommes tout aussi colossales sont prévues pour les propositions de budget des années à venir. En janvier 2019, le Congressional Budget Office (CBO) estimait que le coût total de la modernisation du NC3 sur les dix prochaines années avoisinerait les 77 milliards de dollars[84].

Le NC3 est un maillon fondamental de la chaîne décisionnelle de la dissuasion nucléaire, puisqu'il constitue le dispositif prévenant les autorités politico-militaires de l'imminence d'une attaque adverse et donne la capacité au président des États-Unis de procéder au déclenchement du feu nucléaire. Pour les planificateurs du Pentagone, la nécessité de procéder à une modernisation des systèmes et à un accroissement du niveau d'automation s'avérait urgente. Certes, des modernisations régulières des systèmes d'armes eurent lieu depuis la chute du Mur, mais l'infrastructure électronique du NC3, pour sa part, reposait encore sur des dispositifs hérités de la guerre froide au point de risquer l'obsolescence. Au travers de la proposition de budget déposée par le Department of Defense, ce n'est ni plus ni moins qu'une refonte du NC3 qui est envisagée. Celle-ci doit permettre au président des États-Unis de disposer d'un système d'alerte avancée qui puisse non seulement détecter de façon précise les attaques adverses, mais encore offrir un éventail approprié de réponses dans des conditions de temporalité et de sécurité optimales, notamment dans le cas de frappes nucléaires sur le territoire et d'attaques cybernétiques des infrastructures électroniques[85].

L'architecture NC3 des États-Unis, faut-il le rappeler, fut conçue au siècle passé, à une époque où les menaces qu'il s'agissait de détecter étaient soit le lancement d'ICBM ou le largage de bombes. Les contraintes opérationnelles de telles attaques offraient aux autorités un temps certes court mais néanmoins réel afin de permettre de définir la réponse la plus appropriée à mettre en œuvre. Une fois la menace détectée par les systèmes d'alerte avancée, le président des États-Unis disposait de quelques 30 minutes[86] pour formuler sa réponse, qu'il s'agisse du lancement d'une riposte ou d'un temps de réflexion permettant de faire redescendre le niveau d'alerte en fonction des informations reçues en continu sur l'évolution de la menace détectée. Aujourd'hui, les spécificités nouvelles des armements équipant les arsenaux des principales puissances ont connu des évolutions telles qu'elles

[83] Source : https://www.defense.gov/Newsroom/Releases/Release/Article/2079489/dod-releases-fiscal-year-2021-budget-proposal/.

[84] U.S. Congressional Budget Office, « projected Costs of U.S. Nuclear Forces, 2019 to 2028 », janvier 2019, https://www.cbo.gov/system/files/2019-01/54914-NuclearForces.pdf Voir aussi Michael T. KLARE, « Skynet Revisited: The Dangerous Allure of Nuclear command Automation », *Arms Control Today*, Arms Control Association, avril 2020, *cf*. https://www.armscontrol.org/act/2020-04/features/skynet-revisited-dangerous-allure-nuclear-command-automation.

[85] Michael T. KLARE, *ibid.*, *cf*. https://www.armscontrol.org/act/2020-04/features/skynet-revisited-dangerous-allure-nuclear-command-automation.

[86] 30 minutes dans le cas d'un lancement d'ICBM, 15 minutes dans l'hypothèse du lancement d'un SLBM.

ont considérablement comprimé la temporalité du processus décisionnel en cas de crise. La première spécificité est la proximité entre le plafond de destruction des armements conventionnels et le seuil de destruction des armements nucléaires tactiques. Actuellement, la plupart des missiles balistiques conventionnels sont susceptibles de pouvoir accueillir des charges nucléaires. Les vecteurs disposent par ailleurs d'aptitudes techniques nouvelles telles que la vitesse hypersonique et l'hyper-manœuvrabilité. Ces caractéristiques des vecteurs de nouvelle génération participent donc à une contraction inédite de la temporalité disponible pour la décision de riposte. Du fait de durées de vol limitées à 5 minutes, les systèmes NC3 n'offrent qu'un temps extrêmement limité pour pouvoir déterminer si l'attaque détectée est de nature conventionnelle ou nucléaire. Or, selon la nature de la frappe, la réponse formulée et l'évaluation des risques encourus (notamment en ce qui concerne les retombées) pourra varier de manière considérable. Aux caractéristiques nouvelles des vecteurs de frappe s'ajoute le risque d'attaque cybernétique pouvant affecter la disponibilité et/ou la qualité des informations censées parvenir en temps réel au commandement suprême lors d'une crise ou d'un échange conventionnel/nucléaire. Ces éléments expliquent pourquoi le DoD attache une importance fondamentale à l'intégration d'un certain niveau d'automation au sein des systèmes de commandement et de contrôle. En ce XXI^{e} siècle, les caractéristiques techniques et contextuelles des futures attaques (qu'elle soit de nature nucléaire ou conventionnelle, ou les deux) entraînent deux phénomènes que les états-majors se doivent de résoudre : la *surcharge informationnelle* et *la compression temporelle*. Le principal paradoxe résultant de la numérisation exponentielle de nos systèmes de défense – qui ont recours à des masses de données provenant de multiples capteurs terrestres, aériens, maritimes, spatiaux et cybernétiques – réside dans la production d'une somme de données que ne peuvent traiter des opérateurs humains (du moins, pas à la vitesse de traitement souhaitée). Or la rapidité avec laquelle les attaques futures seront conduites par une puissance adverse mèneront à une réduction inédite du temps de la décision. Aux yeux de certains analystes, un NC3 assisté par une intelligence artificielle permettra de répondre à ces deux contraintes[87]. C'est notamment le propos de Adam Lowther et Curtis McGiffin : dans un article paru sur le site *War on the Rocks*, ils n'ont pas hésité à affirmer que la compression inédite du temps de réaction à une attaque (conventionnelle majeure ou nucléaire) a désormais placé les États-Unis dans une situation telle que les autorités du pays ne seraient plus en mesure de réagir assez rapidement. Durant la dernière décennie, la Russie a procédé à de nombreuses transformations de son arsenal conventionnel et nucléaire. Elle a principalement axé la modernisation de ses systèmes en attachant un soin particulier à rendre ceux-ci aussi indétectables que possible. Parmi les nouveautés développées par la Russie figurent notamment les systèmes de croisière Kaliber-M[88] et Kh-102, le vecteur sous-marin inhabité *Poseidon Ocean Multipurpose System Status-6* ou encore le système hypersonique Avangard Objekt 4202. Tous ces dispositifs ont été conçus afin de contourner les capacités NC3 existantes des États-Unis.

La compression de la boucle OODA (Observation, Orientation, Décision & Action) liée au déploiement de nouveaux vecteurs de frappe par les adversaires des États-Unis s'avérera incompatible avec la préservation d'un créneau décisionnel humain. À défaut d'un tel système, les États-Unis ont, toujours selon les mêmes auteurs, trois possibilités afin d'échapper au défi de la compression de la boucle décisionnelle.

[87] Kelly M. SAYLER, *Artificial Intelligence and National Security*, CRS Report, R45178, 21 novembre 2019.
[88] À titre d'illustration, un missile Kaliber-M tiré en mer Noire a parcouru la distance de 250 km en 137 secondes. Voir https://fr.sputniknews.com/defense/201912111042580263-un-missile-kalibr-frappe-une-cible-en-mer-noire-parcourant-250-km-en-137-secondes--video/.

Illustration 4 : Lancement d'un missile de croisière à longue portée Kaliber-M depuis la frégate Amiral Essen en mer Méditerranée

(source : Service de presse du ministère de la Défense de la Fédération de Russie, via AP)

Une première possibilité afin de permettre aux autorités politico-militaires de disposer d'un temps supplémentaire pour déterminer la réponse à formuler en cas d'attaque serait de recentrer le dispositif nucléaire sur les capacités de seconde frappe afin de leur garantir une plus grande survivabilité en cas d'attaque et d'assurer à l'agresseur une riposte déterminante et hors de toute proportion imaginable avec l'enjeu. Cette option implique pour le pays d'être en mesure d'absorber une première frappe, de consentir à la perte de vies humaines et peut-être à la décapitation d'une partie des organes de décision.

Une seconde option de modernisation consisterait pour l'état-major américain de concentrer ses investissements sur les capacités de surveillance et de reconnaissance, et ce dans le cadre d'une stratégie de préemption. En d'autres termes, la stratégie de dissuasion des États-Unis s'axerait sur la précision de ses systèmes d'alerte avancée afin de parer à tout effet de surprise auquel pourrait recourir un éventuel adversaire. La menace d'une frappe préemptive incapacitante, conduite sur la base d'information d'alerte précoce, suffirait alors à faire renoncer l'adversaire à conduire des manœuvres laissant supposer l'imminence d'une attaque de sa part. Le degré de fiabilité et de précision sur lequel devrait s'appuyer un tel dispositif de surveillance et de reconnaissance exigerait l'assistance d'un système d'IA capable de procéder à la récolte, la fusion et l'analyse de la masse de données en provenance d'une multitude de senseurs. La stratégie du Pentagone se révélerait dépendante d'une intelligence artificielle dont les inférences seraient soit inexplicables, soit incompréhensibles aux organes de décision.

Enfin, une troisième option entre les mains des stratégistes américains – toujours en vue d'échapper aux contraintes de la compression du tempo décisionnel découlant de la nature des moyens de l'adversaire et du recours à l'automation – supposerait que les États-Unis concentrent à leur tour leurs investissements dans des capacités en mesure de compresser le tempo décisionnel de l'adversaire. Il s'agirait, entre autres, de poster des systèmes de frappe à proximité immédiate des frontières terrestres ou navales du ou des adversaires visés. La crainte d'une attaque surprise et la perspective d'une destruction mutuelle assurée conduiraient les protagonistes à converger vers un accord destiné à échapper à une situation dans laquelle la menace d'instabilité ne conviendrait à aucune partie. L'option en question suppose cependant que les protagonistes s'inscrivent dans une culture de négociation commune et perçoivent leurs intérêts de sécurité selon des critères convergents. Il suffit qu'une seule puissance décide de ne pas inscrire sa stratégie de sécurité dans un cadre multilatéral pour qu'une course aux armements déstabilisatrice soit relancée.

Pourtant, une quatrième option existe. Pour Lowther et McGiffin, l'intégrité et l'utilité future d'un système NC3 capable de relever le défi de la compression du tempo décisionnel découlant de la modernisation des armements adverses exige que lui soit adjointe une intelligence artificielle en mesure de procéder aux décisions indispensables en cas de détection d'un lancement d'ICBM ou de SLBM. Pour les analystes, il est donc indispensable que les États-Unis dotent leur NC3 d'une IA capable de détecter, de décider et de diriger les forces stratégiques grâce à des réponses prédéterminées, constamment réévaluées en fonction de données obtenues en temps réel, et ce afin de permettre aux forces stratégiques de répondre et de défaire dans des contraintes de temps extrêmes toute attaque qui serait engagée contre le pays[89]. À quoi pourrait ressembler, dès lors, une intelligence artificielle programmée pour l'assistance des décideurs politico-militaires ? Sans doute, celle-ci présentera-t-elle de nombreuses caractéristiques du programme conduit par la DARPA en vue de permettre à une IA de collecter et d'interpréter les données issues de phénomènes provenant du monde réel, c'est-à-dire découlant de l'ensemble des innombrables interactions résultant des décisions politiques de tout niveau, des dynamiques sociales, des échanges économiques et des variables culturelles[90]. À l'évidence, une telle IA ne pourra reposer sur un apprentissage machine basé sur la production d'itérations multiples. Dans le cadre d'un conflit nucléaire, il n'existe pas d'apprentissage possible par itérations. Il serait donc question d'une intelligence artificielle générale pour appuyer les analyses de données et leur traitement.

On peut toutefois s'inquiéter avec Michael T. Klare de la voie dans laquelle s'orienterait la sécurité internationale dans le cas où une puissance comme les États-Unis en viendrait à confier à des algorithmes évolués (disposant éventuellement de réponses préprogrammées) la conduite de la stratégie nucléaire en temps de crise. Une première inquiétude réside dans la qualité des systèmes d'IA associés au processus décisionnel. Un système d'IA de qualité est avant tout un système d'IA entraîné. Or le choix d'une riposte à formuler en cas d'attaque nucléaire, qu'elle soit suspectée ou avérée, diffère quelque peu de la prévision comportementale d'un consommateur sur la base des données issues d'achats passés ou provenant de consommateurs semblables de par le monde. En matière de dissuasion et de riposte nucléaire, les données réelles sont extrêmement rares : à l'exception des deux bombes atomiques lancées sur Hiroshima et Nagasaki, aucun emploi ultérieur d'armes atomiques ou nucléaires n'a – fort heureusement – eu lieu. Cette absence de données réelles d'échanges nucléaires passés rend donc impossible l'entraînement des systèmes d'IA pour l'élaboration de scénarios en cas de crise. Le matériau sur lequel s'appuient les systèmes d'IA censés

[89] Adam LOWTHER, Curtis MCGIFFIN, « America needs a 'Dead hand' », *War on the Rocks*, 16 août 2019, *cf.* https://warontherocks.com/2019/08/america-needs-a-dead-hand/.
[90] Pour plus de détails, *cf.* https://www.darpa.mil/news-events/2019-01-04.

assister les organes décisionnels politico-militaires n'est donc constitué que de résultats provenant de simulations de combat et de jeux de guerre. Il découle par conséquent de cette situation que les algorithmes qui assisteront les décideurs humains à propos des réponses à formuler dans l'hypothèse d'une attaque engagée par une puissance nucléaire adverse comporteront des biais liés à l'insuffisance de données réelles sur lesquelles ces systèmes d'IA n'auront donc pas pu bâtir leur « expertise ».

Une seconde inquiétude susceptible de venir se greffer aux limitations d'un système d'IA censé appuyer les organes de décision politico-militaires réside dans le biais cognitif pouvant affecter les décideurs humains en raison de la foi aveugle qu'ils accorderaient à une IA nécessairement imparfaite. Une étude menée par la RAND Corporation en 2018 a démontré à quel point un décideur humain avait tendance à considérer l'IA comme son équivalent. Cette confiance injustifiée, précisait l'étude, pouvait être génératrice de risques nouveaux en matière de dissuasion nucléaire[91].

Illustration 5 : le Poseidon Ocean Multipurpose System Status-6 (vision d'artiste)

9. Conclusion partielle

Ce chapitre a démontré la relation étroite qui a toujours existé entre l'intelligence artificielle et la dissuasion nucléaire. Très tôt, chacune des superpuissances de la guerre froide a envisagé de recourir à l'automation en vue d'assurer la fiabilité de la dissuasion, c'est-à-dire la garantie d'une sauvegarde des capacités de seconde frappe. Il pourrait être affirmé que l'IA a très tôt été perçue comme la seule garante d'un semblant de continuité de la décision et de la stratégie dans l'éventualité d'une destruction mutuelle assurée. Les limites des solutions technologiques disponibles à l'époque, sans compter les nombreuses fausses alertes déclenchées par les systèmes de détection avancés, ont poussé les instances décisionnelles à revoir leur opinion du recours à l'automation comme ultime garante de la sécurité. Il n'en demeure pas moins qu'une réflexion a toujours existé, tant et si bien qu'à l'heure actuelle, au vu des percées récentes de l'apprentissage machine, la possibilité de confier à des machines la prise de décision en cas de conflit majeur conventionnel ou nucléaire n'est plus seulement au stade de l'étude. Une question demeure cependant : l'IA constituera-t-elle un facteur d'égalisation de la puissance ou, au contraire, le déterminant d'un fossé infranchissable entre les organisations militaires ?

[91] Edward GEIST, Andrew J. LOHN, (eds), *How Might Artificial Intelligence Affect the Risk of Nuclear War?*, RAND Corporation, Santa Monica (Calif.), Serie Perspective, 2018.

IV. IApocalypses

1. Singularité, suprématie quantique et fin de l'Humanité : faut-il craindre l'IA ?

Depuis de nombreuses années, des scientifiques de haut rang mènent régulièrement une charge contre des applications issues de l'intelligence artificielle. Au cours des dernières décennies, de multiples lettres ouvertes ont été publiées qui tentent de sensibiliser l'opinion publique, ainsi que les décideurs, aux dangers que comporterait l'IA pour le devenir de l'Humanité. Plusieurs éléments sont à souligner à ce propos. Le premier consiste à observer que ces dénonciations sont loin de constituer un phénomène récent. Déjà, en 2000, le fondateur de la société Sun Microsystems, Bill Joy, avait fait paraître dans la revue *Wired* un article, entre-temps devenu célèbre, intitulé « Why The Future Doesn't Need Us » et dans lequel il exposait ses craintes quant aux conséquences d'une liaison trop étroite entre la robotique, les nanotechnologies, les biotechnologies et les sciences cognitives. Paru aux États-Unis, l'article de Bill Joy intervenait alors en plein débat sur les perspectives des nanotechnologies dans le cadre de la convergence technologique et de la National Nanotechnology Initiative (NNI) de la Maison Blanche. Un second élément à prendre en considération est la personnalité des signataires de telles lettres ouvertes. Loin de rassembler un parterre de figures de la défense des droits de l'homme ou d'organisations non gouvernementales, les articles en question regroupent des industriels de renom – souvent liés à des entreprises de pointe et issus de la Sillicon Valley – ou des scientifiques parmi les plus doués de leur génération. Enfin, un troisième aspect de la démarche à la base de ces lettres ouvertes est l'appel lancé par ces entrepreneurs aux États qu'ils jugent désormais chargés de définir les garde-fous face aux technologies d'IA en train d'être développées. Il s'agit, ni plus ni moins, d'un appel des libertariens aux structures gouvernementales en vue de juguler – du moins en apparence – les dérives susceptibles de découler des avancées de l'IA.

a) Bill Joy et l'extinction de l'espèce humaine

Premier cri d'alarme au sein de la communauté entrepreneuriale américaine, l'appel lancé par Bill Joy à travers un article de la revue *Wired* constitua un véritable boulet rouge qui mit un terme brutal à la quiétude du progressisme ambiant des technologies numériques, du moins au sein d'une communauté d'experts avertis. Pour rappel, le début du XXI^{e} siècle constitue un âge d'or de l'Amérique de l'après-guerre froide. Portés par une croissance économique inédite, les États-Unis s'établissent en véritable leader du système international et exercent une domination pratiquement incontestée dans le secteur des technologies de pointe, des ordinateurs et des réseaux. Dans ce nouveau positivisme ambiant, teinté d'un certain scientisme, les propos exprimés par Bill Joy font office d'empêcheur de « développer en rond ». Tout en rappelant que de tout temps les nouvelles technologies furent accompagnées de préoccupations d'ordre éthique, Bill Joy affirme avoir pris conscience de la particularité de la révolution technique du nouveau millénaire à la suite d'une discussion qu'il entretint au sortir d'une conférence avec le philosophe John Searle et Ray Kurzweil, en particulier au sujet de l'ouvrage de ce dernier (alors à paraître) : *The Age of Spiritual Machines*. Au cours de cette conversation, Ray Kurzweil tentait de faire la démonstration selon laquelle les hommes étaient voués à fusionner avec la machine en attendant que cette dernière ne surpasse l'intellect humain. La thèse défendue par Ray Kurzweil fut, plus tard, développée dans son ouvrage dans lequel il évoque le concept de « Law of Accelerating Returns » qui succéderait à la loi de Moore dans l'esprit des scientifiques des premières décennies du développement de l'informatique.

Le propos central de Bill Joy portait sur la différence fondamentale existant entre les précédentes révolutions technologiques et la révolution numérique qui ne faisait alors que commencer à produire ses effets. Le fondateur de Sun Microsystems admettait que chaque révolution technologique générait

dans son sillage des transformations sociales, politiques et économiques parfois majeures sans qu'elles ne viennent pour autant altérer la marche de l'Humanité – si tant est que l'on suppose l'existence d'une telle « marche ». Pour Bill Joy, « les technologies les plus incontournables du XXIe siècle – la robotique, le génie génétique et les nanotechnologies – représentent une menace différente des technologies antérieures. Concrètement, les robots, les organismes génétiquement modifiés et les " nanorobots " ont en commun un facteur démultipliant : ils ont la capacité de s'autoreproduire. Une bombe n'explose qu'une fois ; un robot, en revanche, peut proliférer et rapidement échapper à tout contrôle. » Il poursuit sa réflexion en se demandant quelles seraient les différences pertinentes entre les armes auxquelles aboutiraient les technologies GNR (génétique, nanotechnologies et robotiques) et celles qui ont émergé durant le XXe siècle et plus spécifiquement la dernière guerre mondiale. « Certes, [admet Bill Joy,] les technologies liées aux armes de destruction massive – nucléaires, biologiques et chimiques (NBC) – étaient puissantes, et l'arsenal faisait peser sur nous une menace extrême. Cependant, la fabrication d'engins atomiques supposait, du moins pendant un temps, l'accès à des matériaux rares – et même inaccessibles –, autant qu'à des informations hautement confidentielles. Au surplus, les programmes d'armement biologiques et chimiques exigeaient souvent des activités à grande échelle. Les technologies du XXIe siècle […] sont porteuses d'une puissance telle qu'elles ont la capacité d'engendrer des classes entières d'accidents et d'abus totalement inédits. Circonstance aggravante, pour la première fois, ces accidents et ces abus sont dans une large mesure à la portée d'individus isolés ou de groupes restreints. En effet, ces technologies ne supposent ni l'accès à des installations de grande envergure, ni à des matériaux rares ; la seule condition pour y avoir recours, c'est d'être en possession du savoir requis[92]. »

b) Vers un moratoire ?

La charge menée par Bill Joy a stimulé de nombreux échanges entre les acteurs de la communauté entrepreneuriale américaine à propos des perspectives de développement des technologies GNR durant la première décennie des années 2000. S'opposaient alors les médiatiques partisans d'une vision optimiste des ruptures technologiques, des scientifiques, industriels, philosophes, sociologues convaincus de l'avènement de *transhumanisme*, un âge futur durant lequel l'homme parviendrait à s'affranchir des contraintes physiologiques de sa biologie pour progressivement atteindre une certaine éternité. Cette vision optimiste des ruptures technologiques GNR et de l'IA ne s'encombre point de considérations relatives aux enjeux en matière de sécurité internationale. Face aux techno-évangélistes annonciateurs des heures heureuses du transhumanisme, les partisans d'une vision dystopique se sont fait jour. C'est au travers d'une lettre ouverte publiée dans le journal *The Independent*, le 1er mai 2014, que Stephen Hawking, physicien et cosmologiste britannique de renom, adresse une mise en garde à l'opinion publique et aux décideurs du monde. D'une certaine façon, Stephen Hawking reprend dans une large mesure la substance des propos émis quelque 15 années plus tôt par Bill Joy. Il affirme que les technologies se développent aujourd'hui à un tel rythme qu'elles deviendront très vite incontrôlables au point de placer l'Humanité devant un péril sans précédent. La tribune de Stephen Hawking est co-signée par de prestigieux scientifiques[93] qui rejoignent les craintes de Hawking quant aux dangers auxquels conduiraient les progrès extraordinaires accomplis, notamment, par l'intelligence artificielle. Car, il convient de le préciser, au travers de la dénonciation des « technologies », c'est bien l'intelligence artificielle qui est appelée à la barre par les auteurs de cette tribune. Les cosignataires de la lettre ouverte énoncent quelques exemples des avancées hors du commun réalisées grâce ou avec l'intelligence artificielle (voiture autonome, l'assistant SIRI de la

[92] Bill JOY, « Why The Future Doesn't Need Us », *Wired*, *cf.* https://www.wired.com/2000/04/joy-2/

[93] On citera, entre autres, Max Tegmark, professeur de physique théorique au MIT, Stuart Russell, professeur d'intelligence artificielle à l'Université de Berkeley ainsi que Frank Wilczek, professeur de physique au MIT et prix Nobel de physique.

société Apple, ou encore la victoire remportée par une machine intelligente au jeu Jeopardy face à des concurrents humains, etc.). Ils ajoutent encore que la croissance exponentielle des données de masse obtenues au travers de l'utilisation faite par les individus de leurs outils numériques (ordinateurs, smartphones, réseaux sociaux, logiciels d'assistance médicale, applications de performances sportives, etc.) offrira aux systèmes d'intelligence artificielle une quantité d'informations telle sur les pratiques humaines dans absolument tous les secteurs qu'ils développeront des inductions automatiques imprévisibles. Cette imprédictibilité ne résultera pas du caractère aléatoire des liaisons opérées entre les données, mais bien du fait qu'elles découleront d'une puissance de calcul et de liaison d'une telle complexité que l'intellect de l'homme ne sera plus en mesure d'accéder à la « boîte noire » des systèmes d'IA.

L'affaire ne s'arrête pas là : Stephen Hawking réitère ses attaques. Sur la BBC, l'astrophysicien répète que l'intelligence artificielle pourrait conduire à l'extinction de la race humaine. Elon Musk, dans le sillage de Hawking, fait une nouvelle fois part de ses inquiétudes en soulignant le danger que l'intelligence artificielle pourrait faire peser sur l'homme. Bill Gates, co-fondateur de Microsoft, évoque son pessimisme quant aux perspectives de l'IA lors d'une allocution à l'occasion de la conférence AMA le 28 janvier 2015. C'est ce même mois de l'année 2015 qu'une lettre ouverte co-signée par une multitude de chercheurs en IA paraît sur le site Future of Life Institute[94]. Le document en question expose, avec précision, le raisonnement conduisant les signataires à avertir l'opinion publique des dangers auxquels exposera demain, à moyen terme, l'IA :

> « *The adoption of probabilistic and decision-theoretic representations and statistical learning methods has led to a large degree of integration and cross-fertilization among AI, machine learning, statistics, control theory, neuroscience, and other fields. The establishment of shared theoretical frameworks, combined with the availability of data and processing power, has yielded remarkable successes in various component tasks such as speech recognition, image classification, autonomous vehicles, machine translation, legged locomotion, and question-answering systems.* »

Toutefois, les signataires de la lettre ouverte du Future of Life Institute soulignent un autre aspect des recherches menées dans le domaine de l'IA qui, selon eux, pourraient conduire à une « perte de contrôle » de ce qui est issu de la recherche et développement. C'est, plus spécifiquement, la liaison désormais établie entre les laboratoires et les perspectives commerciales de l'IA qui forment la base des inquiétudes des signataires. Les solutions d'IA conçues au sein des laboratoires donnent désormais lieu à des produits commercialement rentables dont les bénéfices sont réinjectés dans la recherche de solutions d'IA encore plus performantes. Ce cycle aboutit à des fertilisations croisées entre de nombreux secteurs d'activités et domaines de recherche investis principalement par des fonds issus du secteur privé. Aujourd'hui, il n'est pas un aspect de la vie humaine qui ne soit pas concerné par une solution d'IA. Pour les auteurs de la lettre ouverte, il est plus que temps de nous focaliser sur les apports vertueux des domaines d'application de l'IA et non uniquement sur les moyens de rendre les systèmes d'IA plus performants (un objectif qui est aujourd'hui largement atteint par les cycles de réinvestissements des bénéfices économiques résultant de l'IA). La lettre ouverte insiste donc sur la nécessité de définir des orientations résolument nouvelles pour la recherche afin de garantir une meilleure capacité de résistance des systèmes, à leur vérification et à leur profitabilité (en termes autres qu'économiques) pour l'ensemble de la société[95].

[94] Cette lettre ouverte est disponible à la lecture depuis l'adresse suivante : https://futureoflife.org/ai-open-letter

[95] Jean-Gabriel GANASCIA, *Le mythe de la Singularité : faut-il craindre l'intelligence artificielle ?*, Paris, Seuil, coll. Points, 2017, p. 11.

La lettre ouverte du Future of Life Institute constitue une des nombreuses initiatives émanant de divers groupes de recherche et de sensibilisation sur les questions relatives aux technologies émergentes, en premier lieu desquelles figure le domaine de l'intelligence artificielle. On pourrait, du reste, citer quelques-uns de ces acteurs institutionnels : Extropy, la Singularity University, l'Institute for Ethics and Emerging Technologies (IEET), le Machine Intelligence Research Institute (MIRI) ou encore le Centre for the Study of Existential Risk (CSER, basé à l'Université de Cambridge en Angleterre). Il existe un foisonnement évident de ce type d'entreprises à travers le monde, pas seulement dans le monde anglo-saxon, mais aussi en Europe[96]. Nombre de ces initiatives portées par des institutions ou des *think tanks* sont financées par les représentants de l'industrie du numérique, quand elles ne sont pas directement fondées par des entreprises du secteur des TCI telles que Google, Cisco, Nokia, Autodesk ou Genentech. Ces organismes et leurs messages, tous d'une grande diversité, ont néanmoins un point commun : ils annoncent un événement brutal dont les conséquences pour l'Humanité s'avéreront tout aussi radicales qu'irréversibles, à savoir l'avènement de la *singularité*. Pour certains, la *singularité* est pleine de promesses. Pour d'autres, elle constitue un risque existentiel majeur pour l'avenir de l'Humanité. Cette dernière posture suppose, entre autres dangers, la perte de contrôle sur des machines dont le degré d'intelligence serait parvenu à dépasser celui de l'homme ou la possibilité d'une dépendance telle de l'homme aux machines intelligentes que toute décision humaine serait, d'une manière ou d'une autre, tributaire d'une IA.

La *singularité* constituerait donc l'événement majeur qui bouleverserait l'ordre mondial dans sa totalité, qu'il s'agisse des relations économiques, des rapports diplomatiques, des équilibres militaires, de la condition sociale et humaine, des conceptions philosophiques portant sur l'homme et la nature. Mais qu'entend-on exactement au travers de cette *singularité* présentée tout à la fois comme le pire des maux et le meilleur des remèdes à tous les problèmes de l'Humanité ?

On rappellera, tout d'abord, que la crainte de la survenance d'un événement tout aussi brutal que disruptif ne constitue pas une nouveauté dans l'histoire récente de l'Occident. S'agissant du rapport homme-machine et de ses implications sécuritaires au sens large (c'est-à-dire susceptibles d'emporter le sort de l'Humanité), des scénarios ont foisonné depuis de nombre d'années. Dès 1962, l'un des statisticiens ayant œuvré aux côtés d'Alan Turing durant la Seconde Guerre mondiale, Irvin John Good, avait fait mention de « spéculations » relatives à la possibilité de développer dans le futur une machine intelligente[97]. Pour Good, la possibilité d'une explosion de l'intelligence dans les années à venir semblait constituer un scénario inéluctable tout en précisant qu'une telle explosion ne pourrait passer que par la machine[98]. L'évocation de la *singularité* émane pour la première fois d'un roman de science-fiction (comme souvent) écrit par Vernor Vinge. Il faut toutefois attendre les années 1980 pour que celui-ci le théorise au sein d'un essai intitulé *The Coming Technological Singularity*, paru en 1993. Le propos de Vinge consiste à affirmer qu'au terme de moins de trente ans, les avancées qui surviendront dans les technologies de l'information et des communications aboutiront à doter les machines d'une intelligence surhumaine, reléguant l'homme à un statut secondaire. À moins que l'homme ne puisse profiter des « bienfaits » de la suprématie de la machine et ne s'élève au niveau de

[96] On observera, au demeurant, que le courant transhumaniste trouve ses racines en Europe et non dans le monde anglo-saxon. Voir à ce propos Béatrice JOUSSET-COUTURIER, *Le transhumanisme : faut-il avoir peur de l'avenir ?*, Paris, Eyrolles, 2016.

[97] L'intitulé exact de son allocution était « Speculations concerning the first ultraintelligent machine ».

[98] L'après-Seconde Guerre mondiale assiste au foisonnement de nombreux scénarios et de multiples interrogations sur le devenir des rapports entre l'homme et la machine, et ce dans un contexte particulier lié à la course scientifique, technologique, industrielle et militaire que se livrent alors les États-Unis et l'Union soviétique sur fond de lutte idéologique. Dès 1950, le mathématicien Stanislaw Ulam émet l'idée d'une discontinuité à venir du fait des progrès de la technologie. Dans son sillage, Isaac Asimov, célèbre auteur de science-fiction notamment connu pour sa série de romans sur les robots, fait paraître en 1956 une nouvelle intitulée *The Last Question,* dans laquelle il envisage le développement d'un ordinateur à la dimension de l'univers parvenant à renverser la seconde loi de la thermodynamique en faisant décroître l'entropie.

celle-ci. L'évolution des rapports entre le combattant et la machine depuis l'avènement de l'ère industrielle, surtout depuis la rupture technologique survenue lors de la Seconde Guerre mondiale, n'est pas étrangère à la conception de l'évolution humaine que dépeint Vernor Vinge. Au cours de l'Histoire, la question de l'évolution du combattant a principalement consisté à adapter la machine à ce dernier pour qu'il puisse en extraire, lors du combat, les meilleurs résultats. L'ergonomie militaire impliquait donc une conformation de la machine – c'est-à-dire l'armement – à l'homme – c'est-à-dire le soldat. Dans le cadre de la compétition que se sont livrée les États-Unis et l'Union soviétique dans le secteur spatial, le défi pour les ingénieurs avait tout autant consisté à étudier la conception d'une machine susceptible de porter l'homme dans un environnement aussi hostile que l'espace que de pousser l'homme lui-même dans les retranchements de son humanité pour qu'il s'adapte au mieux aux contraintes de cet environnement. Avec la révolution des ordinateurs et des réseaux qui émerge dans le courant des années 1970, la perspective d'une ergonomie inversée entre l'homme et la machine est imaginée, l'idée étant désormais d'amener l'homme à s'adapter aux exigences de la machine pour garantir la meilleure performance de celle-ci au combat. L'idée sous-jacente au propos de Vernor Vinge s'inscrit donc, pourrait-on dire, dans l'ère du temps : Vernor Vinge verbalise ainsi une approche sur laquelle les planificateurs militaires se penchent de longue date. Les différentes visions et extrapolations qui émanent de scientifiques et technologues investis dans le champ de l'IA ne formulent cependant aucune réponse à une question fondamentale : l'homme demeurera-t-il encore demain en mesure de comprendre une machine dont l'intelligence aura gagné en complexité ?

2. Finitum Non Capax Infiniti[99]

Les divers scénarios que nous venons d'explorer quant aux répercussions d'une dépendance de nos systèmes de défense à des systèmes d'IA avancés mettaient en exergue le risque d'une perte de contrôle par l'homme des mécanismes de décision stratégique. Le dénominateur commun des différentes conjectures évoquées était donc l'impossibilité pour l'homme de parvenir à conserver ou à récupérer le contrôle d'un système d'IA au cœur de configurations critiques (déploiement de forces, décision de lancement de missiles, etc.). Pourtant, ces différentes hypothèses ne faisaient qu'effleurer le cœur même du problème des rapports à venir entre l'homme et l'intelligence artificielle, à savoir la capacité d'un système d'IA à se rendre intelligible et accessible à l'homme et l'aptitude présente et à venir de l'homme à comprendre un système d'IA avancé. Il s'agit, en réalité, de deux aspects des rapports entre l'Homme et l'IA : l'*inexplicabilité* et *l'incompréhensibilité*. Norbert Wiener, père fondateur de la cybernétique, exprimait déjà la substance de ce problème en 1959 :

> « *[…] les machines préexistantes ne nécessitaient pas que nous interprétions ce qu'elles allaient faire. Les conséquences mécaniques de leur utilisation étaient plus ou moins évidentes, quoique les conséquences sociales ne l'étaient pas. Aujourd'hui, pas plus les conséquences mécaniques que les conséquences sociales ne sont pleinement prévisibles.*[100] »

L'un des risques majeurs que pourrait courir l'Humanité dans les décennies à venir, au fur et à mesure des percées réalisées dans le champ de l'intelligence artificielle, serait de ne plus être en mesure d'évaluer la qualité des processus qui se situent au cœur des fonctions d'une IA. Or, *l'explicabilité* et la *compréhensibilité* des mécanismes se situant au cœur des systèmes d'IA déployés dans un large éventail de secteurs d'activité humaine constituent des conditions indispensables à la bonne intégration de l'IA dans la société. En effet, quels qu'ils soient, et quels que soient leurs domaines

[99] Ce qui est limité ne peut embrasser l'infini.

[100] Norbert WIENER, « Man and the Machine », (interview with Norbert Wiener), *Challenge: The Magazine of Economic Affairs*, No. 7, 1959, p. 37.

d'activité d'appartenance, les utilisateurs de systèmes d'IA ont besoin de comprendre comment les décisions susceptibles de les impacter sont adoptées. En philosophie, il existe effectivement une différence fondamentale entre l'explication et la compréhension. Dans le *Phédon* de Platon, Socrate explique pourquoi il reste là à dialoguer avec ses amis tandis qu'il est sur le point de mourir. Dans son discours, Socrate évoque l'*explication* en indiquant faire reposer celle-ci sur les causes physiques, simples, causales d'un comportement. Il parle ensuite de la *compréhension* qui, pour sa part, nécessite d'entendre les raisons et les valeurs des actes. Expliquer, affirme Socrate, c'est amener à comprendre. Il s'agit là de deux moyens d'action qui ont pour but d'engendrer l'adhésion ou au contraire de conduire au rejet ou à la contestation.

Roman V. Yampolskiy s'est penché sur chacune de ces notions et tente de fournir des réponses à cette question centrale : la finitude humaine est-elle en mesure de comprendre le caractère infini (en apparence) de l'IA ?[101]

Pendant des décennies, les projets d'intelligence artificielle – essentiellement des systèmes experts – ont toujours dépendu de l'expertise humaine et, plus spécifiquement, des compétences des ingénieurs. Cette « liaison » entre ingénieurs et systèmes d'IA permettait de faire en sorte que les mécanismes se situant au cœur des systèmes experts demeurent à la portée de l'entendement humain. Il s'agissait là d'une condition essentielle au maintien des machines et à la sécurité de leur mise en œuvre. Depuis une vingtaine d'années, l'apprentissage machine fondé sur les réseaux de neurones profonds (Deep Neural Networks ou DNN) ont considérablement complexifié la nature des rapports entre l'homme et la machine. Les systèmes d'IA qui reposent aujourd'hui principalement sur de tels réseaux de neurones profonds se révèlent d'une telle complexité qu'ils ne permettent plus d'être compris par un ingénieur humain. En effet, la puissance de calcul des systèmes d'IA combinée aux masses de données sans précédent issues de l'activité numérique des individus à travers le monde se dispense de l'intervention humaine pour le fonctionnement des systèmes d'IA. Or, l'acceptation sociale des systèmes d'IA exige que ceux-ci rendent intelligibles leurs processus internes de décision. Cette préoccupation a conduit une institution comme la DARPA, l'Agence des projets de recherche avancée du Département américain de la Défense, à lancer un programme spécifiquement consacré à l'amélioration de l'intelligibilité de l'IA par l'homme. C'est l'approche adoptée par le programme DARPA XAI (eXplainable Artificial Intelligence). D'une manière générale, le concept de XAI désigne un ensemble de méthodes et de techniques ayant pour but de permettre à l'homme d'accéder aux mécanismes profonds qui se situent à la base des décisions et solutions sur lesquelles débouche un système d'IA. Il s'agit, en d'autres termes, de rendre l'IA intelligible aux yeux d'experts humains. Les avancées en matière d'IA aboutiront à la production de systèmes autonomes qui seront capables de percevoir, d'apprendre, de décider et d'agir par eux-mêmes. Cependant, l'efficacité de tels systèmes sera confrontée à leur aptitude – limitée – à expliquer leurs décisions et actions aux utilisateurs finaux (l'homme). Le problème qu'il s'agit de surmonter dans le cadre de telles méthodes consiste à concilier l'impératif de précision (accuracy) des solutions conçues par l'IA et son interprétabilité. La précision des solutions déployées par un système d'IA dépend de méthodes de prédiction complexes. Or des fonctions simples et interprétables n'aboutissent pas aux meilleurs systèmes prédictifs[102]. Certains algorithmes s'avèrent plus compréhensibles que d'autres, et chaque algorithme s'appuie sur un équilibre entre précision et interprétabilité : les intelligences artificielles bâties sur l'apprentissage machine affichant les niveaux de précision les plus élevés sont presque toujours les moins interprétables. À l'inverse, les modèles d'IA et d'apprentissage machine dont les processus s'avèrent

[101] Roman V. YAMPOLSKIY, « Unexplainability and Incomprehensibility of Artificial Intelligence », *cf.* https://arxiv.org/ftp/arxiv/papers/1907/1907.03869.pdf

[102] L. BREIMAN, « Statistical Modeling : The Two Cultures », *Statistical Science*, 2001, Vol. 16, No. 3, pp. 199-231.

à la portée de la compréhension humaine se révèlent les moins précis. Ce que vise un programme comme DARPA XAI consiste à permettre la production par l'IA d'une « boîte de verre » (*glass box*, par opposition à la *black box* de la cybernétique) qu'un humain situé dans la boucle décisionnelle (human-in-the-loop) – généralement un expert – sera capable de déchiffrer. De la sorte, l'homme sera en mesure :

1. de connaître les raisons pour lesquelles le système d'IA a adopté telle décision et celle-là seule ;
2. d'évaluer les éléments qui permettent de conclure au succès ou à l'échec de la décision ;
3. de juger du degré de confiance qu'il peut accorder au système d'IA ;
4. de prendre les mesures correctives nécessaires pour l'amélioration du processus d'apprentissage machine du système.

On le voit, un programme tel que DARPA XAI ne fait qu'apporter un correctif au système d'IA afin de permettre à l'expert humain de disposer, à défaut d'une compréhension totale des processus ayant conduit l'IA à une décision donnée, d'une capacité d'interprétation des principaux éléments permettant de comprendre le résultat des inférences opérées par les algorithmes. En d'autres termes, l'appréhension exhaustive de la « boîte noire » se situant au cœur d'une IA n'est pas nécessaire afin de comprendre les principales motivations ayant conduit à la décision. De la même façon qu'un système d'IA prendra en compte plusieurs centaines de variables avant d'accorder un prêt financier à un individu, les ressorts de la décision pourront sans doute se limiter à 3 ou 4 éléments à la portée des capacités de calcul d'un expert humain. La connaissance des centaines de paramètres ne modifierait en rien la décision finale consistant à accorder ou à refuser le prêt.

Toutefois, au-delà du caractère superflu de certaines variables entrant en considération pour l'élaboration d'une décision, c'est la capacité intellectuelle de l'expert humain qui est aujourd'hui défiée par les systèmes d'IA les plus avancés, ceux reposant sur des réseaux de neurones profonds. Demain, les machines apprenantes basées sur des DNN verront leurs aptitudes s'accroître selon une progression géométrique tout simplement hors de portée de l'intellect humain. De tels systèmes entraînent leurs algorithmes sur des données de masse desquelles des millions de fonctionnalités vectorielles sont extraites pour appuyer des décisions. Quand bien même un système d'IA de ce type parviendrait à rendre « explicables » les processus sur lesquels il se fonde, ceux-ci seraient tout simplement sans la moindre valeur sémantique pour l'homme. Sans doute existera-t-il une fracture irréductible entre les aptitudes des superintelligences de type DNN et le cerveau humain. De tels systèmes technologiques apparaîtront à l'homme, limité par sa finitude, comme disposant de capacités infinies. Or ce qui est fini ne peut embrasser l'infini.

a) L'explicabilité comme enjeu de société

Comme nous pouvons le constater, l'explicabilité des systèmes techniques et, notamment, des décisions automatisées représente un enjeu éthique colossal de l'automatisation dans laquelle se projettent de manière croissante nos sociétés. Cette question de l'explicabilité – qu'il faut comprendre tout à la fois comme l'intelligibilité (autrement dit, la réponse à la question « comment ça marche ? ») et l'éthique de responsabilité (c'est-à-dire la réponse à la question « qui est responsable de la façon dont ça marche ? ») – est la modalité qui permettra l'application d'une IA éthique et juste[103]. La question de l'explicabilité représente donc l'enjeu technique fondamental des systèmes d'intelligence artificielle et c'est elle dont dépendra l'acceptabilité sociale de l'IA. Or, pour quelques experts, la quête d'une intelligence artificielle pouvant expliquer les processus qui se situent à la base

[103] Luciano FLORIDI, Josh COWLS, « A Unified Framework of Five Principles for AI in Society », *Harvard Data Science Review*, juillet 2019, *cf.* https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/6

de sa décision est un objectif inatteignable[104]. On peut, en effet, concevoir un système d'IA capable de construire une fausse explication dont le seul objectif serait de satisfaire l'entité désireuse d'obtenir des informations relatives au processus décisionnel. Ce phénomène serait de nature à rendre inatteignable l'objectif visant le développement d'une IA « responsable ». À moins de tester à plusieurs reprises l'IA en lui soumettant deux fois les mêmes paramètres d'une requête en explication. Si la réponse de l'IA venait à varier au fil des requêtes, il serait fort probable que celle-ci ne pourrait être qualifiée de « responsable ». À l'inverse, la régularité des réponses tendrait à confirmer que l'on se trouve bel et bien en face d'une IA éthique. Toutefois, un tel processus de vérification serait, dans la réalité, quasi impossible à mettre en œuvre. En effet, il serait extrêmement difficile d'imaginer pouvoir introduire plusieurs requêtes en espérant détecter les incohérences sur seulement deux informations parmi les milliers de variables que traitent les systèmes d'IA.

Le contrôle des systèmes d'IA et de leur caractère éthique exige, par ailleurs, la création d'organismes de vérification indépendants qui puissent inspecter non seulement la qualité de la réponse produite par ces systèmes, mais s'investir dans ce que l'on désigne communément par la « boîte noire » du système. À cette exigence, deux obstacles se présentent. Le premier – à supposer qu'une telle instance de vérification puisse voir le jour – consiste à disposer d'experts à même de comprendre les processus internes d'un système d'IA opérant à l'aide de réseaux de neurones profonds (*Deep Learning*). L'évolution des systèmes d'IA fondés sur l'auto-apprentissage engendrera une raréfaction du nombre d'experts humains qui seront encore en mesure de comprendre les processus internes des algorithmes avancés. Il sera donc de plus en plus difficiles pour des instances de vérification indépendants de recruter des experts capables de décrypter de tels algorithmes. Le second obstacle pourrait découler du conflit susceptible d'apparaître entre la préservation par le concepteur des secrets de fabrication d'un système d'IA et la nécessité de procéder à la vérification de la rectitude des processus algorithmiques à la base des réponses formulées par l'IA. En Allemagne, la question d'une *police algorithmique* a déjà été posée. Toutefois, l'envergure de ses missions reste imprécise[105].

Il importe de reconnaître que l'une des difficultés principales est de définir ce que l'on entend par explicabilité. Ce concept est, par « nature », polysémique et recouvre une large gamme d'interprétations allant de « l'explicabilité procédurale » – Comment faire, concrètement ? Quelles formalités accomplir ? Quelle succession d'opérations à exécuter ? – à la « compréhension des critères » – c'est-à-dire la détermination des éléments pris en considération et leur pondération les uns par rapport aux autres. Enfin, on peut encore comprendre la notion d'*explicabilité* en y voyant une clarification des enjeux et des motivations qui se situent à la base d'une décision.

Un enjeu fondamental réside également dans l'interfaçage de l'explication donnée par le système d'IA. Autrement dit, c'est une chose de donner au système d'IA la capacité de clarifier le processus interne qui l'a conduit à l'adoption sans interférence humaine d'une décision donnée, c'en est une autre de permettre à ce système de traduire cette explication à travers une interface accessible soit au vérificateur humain soit à un autre objet technique qui agirait comme médiateur (un peu à l'exemple du programme XAI de la DARPA). La question de l'interfaçage de l'explication affecte inévitablement l'*auditabilité* du système (c'est-à-dire la possibilité qu'offre le système de se rendre « examinable » par un tiers).

[104] Erwan Le Merrer, Gilles Trédan, « The Bouncer Problem: Challenges to Remote Explainability », 3 octobre 2019, *cf.* https://arxiv.org/pdf/1910.01432v1.pdf

[105] Nicolas Kayser-Bril, « 'Explainable AI' Doesn't Work for Online Services – Now There's Proof », 12 novembre 2019, *cf.* https://algorithmwatch.org/en/story/explainable-ai-doesnt-work-for-online-services-now-theres-proof

L'explicabilité et la compréhensibilité des décisions adoptées par un système d'IA constituent une obligation légale et morale. Obligation légale car il en va de la confiance des « calculés » à l'égard des systèmes de calcul avancé (tout comme une relation de confiance doit s'établir entre les administrés et l'administration en permettant aux premiers de disposer des décisions de leur administration et, si le cas se présente, de contester ces décisions sur la base des motivations exposées). C'est d'ailleurs en ce sens que la RGPD adoptée au niveau de l'Union européenne évoque le « droit d'explication » des décisions automatisées.

b) La dimension militaire de la question de l'explicabilité et de la compréhensibilité

On perçoit intuitivement la portée de la question de l'explicabilité et de la compréhensibilité dans le champ de l'action militaire. Demain, lorsque les référents des soldats en opération seront constitués de systèmes d'assistance à base d'intelligence artificielle pour la supervision de la conduite tactique, la parfaite compréhension par le soldat des critères pris en considération par la machine pour la production d'une orientation ou d'une décision sera une condition *sine qua non* à la légalité des actions du soldat. C'est au niveau du rapport entre le soldat humain et la machine que la tension existant entre, d'une part, le devoir d'obéissance du militaire et, d'autre part, sa responsabilité de contester des ordres manifestement illégaux ou criminels atteint son paroxysme. Depuis le procès de Nuremberg, le droit de la guerre a confirmé la norme selon laquelle tout subordonné se doit, dans certaines circonstances, de contester un ordre qu'il juge illégal alors même que sa « culture militaire » l'oblige par principe à une obéissance sans faille. Cette contradiction qui habite tout militaire, liée en grande partie à l'évolution du droit international et des conflits qui ont émaillé le XX^{e} siècle, a donné lieu aujourd'hui à ce que l'on désigne par l'expression de « baïonnette intelligente » : on attend du militaire qu'il incarne tout à la fois le maillon docile et obéissant d'une chaîne de commandement complexe et d'être un excellent juge capable de réflexion personnelle, d'initiative individuelle et d'audace[106].

Les débats récents portant sur la compatibilité du droit de la guerre avec les transformations technologiques découlant des algorithmes avancés (c'est-à-dire basés sur les réseaux de neurones profonds) ne se sont pas attardés sur le véritable problème qui se posera aux militaires de demain. Trop souvent concentrés sur les risques que feraient peser les « systèmes d'armes autonomes » ou sur ce qu'ils considèrent comme tels (puisque de tels systèmes n'existent pas et relèvent seulement de l'automatisme), les détracteurs des technologies émergentes manquent, ni plus ni moins, le défi véritable qui affectera, demain, la conduite des opérations militaires et, plus généralement, de la guerre. Tant que les intelligences artificielles qui seront employées pour assister les états-majors et les chancelleries dans la prise de décision relèveront d'une fabrication et d'une programmation humaine directe, toute erreur d'analyse ou de formulation d'une décision erronée par la machine pourra trouver un responsable humain, qu'il s'agisse du militaire exécutant n'ayant pas repéré l'illégalité de l'ordre ou du programmeur qui aurait, volontairement ou par inadvertance, inculqué à la machine la formulation d'une décision aux conséquences criminelles. Dans un tel cas de figure, l'application du droit – et notamment du droit de la guerre – sera préservée.

[106] Céline BRYON-PORTET, « Du devoir de soumission au devoir de désobéissance ? Le dilemme militaire », *Res Militaris*, *cf.* http://resmilitaris.net/ressources/10123/66/5_res_militaris_article_bryon-portet_texte_inte_gral.pdf

Il en ira tout autrement dans le cas où le système d'IA qui assistera la prise de décision résultera lui-même de la programmation d'un autre système d'IA[107]. Il deviendra autrement plus ardu pour le soldat de disposer d'une explication des processus algorithmiques qui auront présidé à sa décision. Par ailleurs, le rythme des opérations militaires – dont on devine le niveau d'accélération qu'elles connaîtront à l'avenir – ne permettra plus au militaire, et a fortiori au soldat immergé dans l'espace de bataille, de disposer d'une explication accessible, claire et assurément véridique des mécanismes internes de l'IA qui auront guidé son action sur le terrain. L'attribution des responsabilités selon le travail des concepteurs ou les algorithmes des systèmes d'IA ne suffira pas à définir avec certitude les degrés de responsabilité des différents intervenants dans la chaîne de décision. Ce n'est pas seulement l'explicabilité des décisions de l'IA, mais la possibilité même de l'accès à cette explicabilité qui représentera un défi colossal en matière de droit de la guerre. Loin de faciliter la compréhension par le soldat des enjeux qui se situent derrière les ordres reçus, l'adjonction de systèmes d'IA aux mécanismes de décision politico-militaires pourrait ajouter encore au brouillard qui affecte la conduite des forces armées en temps de guerre. On sait à quel point l'appréciation et l'interprétation des ordres peut constituer une opération délicate et risquée lorsque ceux-ci sont soumis au feu de l'action. Or, même quand la règle a bien été enseignée et assimilée, le caractère complexe ou incertain de la situation de fait ne permet pas toujours à l'exécutant (le soldat) d'apprécier si celle-ci tombe sous le coup de la règle. En d'autres termes, le doute peut s'installer quant à l'applicabilité d'une règle claire (du moins en apparence) à une situation qui ne l'est pas[108].

Les enjeux juridiques liés aux limites de l'explicabilité des systèmes d'IA dans le cas d'assistance aux orientations et décisions militaires sont considérables et on ne peut que regretter les raccourcis qu'empruntent les quelques réflexions qui ont tenté d'approcher la problématique. L'adjonction de systèmes d'IA en appui des décisions politico-militaires est loin de constituer une garantie de « progrès » dans la conduite de la guerre. L'IA, contrairement à ce que l'on pourrait penser, ne constitue pas un gage de clarté, pas plus qu'elle ne représente un facilitateur de mise en œuvre de la décision. Face aux incertitudes qui risquent de peser sur le soldat ou, plus généralement, sur tout exécutant militaire de quelque niveau hiérarchique que ce soit, il est à craindre que le droit cède du terrain face à l'ampleur des défis que poseront demain les technologies de l'IA. Une attitude pouvant ici refaire surface serait celle qui consisterait à faire bénéficier d'une immunité pénale généralisée les exécutants militaires appelés à obéir aux consignes que viendraient à leur délivrer à l'avenir des systèmes d'IA en temps d'opération. L'exécutant militaire, compte tenu de l'impossibilité qui sera la sienne de se garantir l'explicabilité des injonctions qui lui sont adressées, ne pourrait être tenu pour responsables des éventuelles exactions commises en cas d'obéissance à un ordre.

3. Superintelligences et superbourdes

Les questions de l'explicabilité et de la compréhensibilité de l'IA s'avèrent fondamentales et exigent de trouver une réponse dans la perspective de l'émergence des machines apprenantes basées sur le DNN et, à plus longue échéance, dans l'éventualité de l'IAG. Si les systèmes d'intelligence artificielle représentent, selon certains, la promesse de résolutions des nombreux problèmes accablant l'Humanité, ils constituent pour d'autres des systèmes technologiques susceptibles – comme d'autres

[107] *Cf.* L'intelligence artificielle de Google a créé une IA plus intelligente que celle créée par l'homme (presse-citron.net). En mai 2017, des chercheurs de Google Brain ont annoncé la création d'AutoML, une IA capable de générer ses propres IA. C'est ce que fit une IA de Google en mettant au point un système de vision par ordinateur surpassant largement les meilleurs dispositifs existants. L'IA créée pour l'occasion par l'IA de Google devrait permettre aux véhicules autonomes et aux robots de disposer d'un mode de vision beaucoup plus performant.

[108] Jacques VERHAEGEN, « Le refus d'obéissance aux ordres manifestement criminels : pour une procédure accessible aux subordonnés », *Revue internationale de la Croix-Rouge*, volume 84, numéro 845, mars 2002, pp. 35-50.

avant eux, et à leur image ou non – d'être en proie à des erreurs ou des dysfonctionnements. La spécificité d'une technologie telle que l'IA ne ne la prémunit en rien contre la survenance de bugs ou de dérapages. Encore faut-il comprendre en quoi de tels dysfonctionnements consistent et comment il peut être possible de les corriger, à défaut de les empêcher.

Toute intelligence artificielle, quel que soit son degré de complexité et de performance, constitue une métastructure technique et, en tant que telle, peut être sujette à des erreurs. L'histoire des technologies regorge d'exemples de systèmes ayant abouti à des résultats imparfaits, des dysfonctionnements ou des crashes. Les exemples cités dans le cadre de notre examen des systèmes d'automation mis en place au sein des dispositifs nucléaires américains et soviétiques durant la guerre froide nous ont rappelé la toute relative fiabilité des systèmes d'IA en matière de surveillance et d'alerte avancée.

La rupture d'un système constitue donc l'une des possibilités du système. Cet enseignement a été popularisé par un adage désormais bien connu de la programmation informatique : « *It's not a bug : it's a feature* » (qu'on pourrait traduire par « ce n'est pas un défaut, c'est une caractéristique »). Cette célèbre expression ironique rappelle pour qui l'aurait oublié que tout système technique programmé, fut-il complexe (et surtout s'il est complexe) est susceptible de connaître des anomalies. Et lorsque ces anomalies s'avèrent impossibles à corriger, elles deviennent finalement une fonctionnalité pervertie du système[109]. Roman V. Yampolskiy énumère trois grandes catégories de « dysfonctionnement » des systèmes d'IA. La première catégorie, que l'on pourrait qualifier de « volontaire », regroupe les actions de programmation délibérées visant à conduire une IA à produire une « décision » aux conséquences néfastes (pour l'Humanité). Une seconde catégorie de dysfonctionnements rassemble les erreurs résultant d'une programmation défectueuse, d'un défaut d'ingénierie ou d'une mise en œuvre prématurée du système. Enfin, une troisième et dernière catégorie comprend les dysfonctionnements résultant de causes extérieures, autrement dit « environnementales »[110].

Si les accidents technologiques ont de tout temps existé, ceux qui sont causés par la défaillance de systèmes d'IA relèvent d'une autre catégorie dans la mesure où ils dépendent de l'« intelligence » contenue au sein même des systèmes à la source de ces défaillances. On ne peut, en effet, mettre sur un pied d'égalité l'accident provoqué par une anomalie d'une machine-outil de type « mécanique » et celle découlant d'une erreur de ligne de code parmi plusieurs centaines de millions. L'actualité regorge d'erreurs rencontrées par des systèmes d'intelligence artificielle de tous les types. En réalité, ces erreurs existent depuis l'émergence même de l'intelligence artificielle entendue ici comme champ disciplinaire. Si les conséquences des premiers incidents rencontrés par les machines développées par les pionniers de l'IA se sont avérés bénins, voire sans la moindre conséquence dans la réalité, les conséquences résultant de telles anomalies se sont, au fur et à mesure des années et du degré d'avancement des machines, avérées plus complexes et socialement plus délicates. On peut ainsi citer les cas d'accident de voitures autonomes (2016), de systèmes prédictifs employés par la police ciblant prioritairement des personnes issues de minorités (2016), celui d'un assistant connecté (Alexa) qui amorce la lecture d'un contenu pour adulte en lieu et place de chansons pour enfants (2017), des logiciels de retouche de portraits photos qui, pour améliorer le visage photographié, décident de blanchir la peau de personnes de couleur noire (2017), un logiciel de Google qui, après une phase d'apprentissage sur les réseaux sociaux, devient homophobe et antisémite ou encore le système de

[109] Nicholas CARR, « It's Not a Bug, It's a Feature. Trite – or Just Right? », *Wired*, *cf.* https://www.wired.com/story/its-not-a-bug-its-a-feature Cette formule signifie qu'une anomalie peut être transformée en fonctionnalité dès l'instant où elle fait l'objet d'une documentation (de telle sorte que personne ne puisse venir s'en plaindre dans la mesure où elle a été spécifiée) ou qu'elle est déclarée comme mineure, voire sans effets sur l'ensemble des fonctionnalités du système.

[110] Roman V. YAMPOLSKIY, « Predicting Future AI Failures from Historic Examples », *Foresight*, novembre 2018, https://www.researchgate.net/publication/329225671_Predicting_future_AI_failures_from_historic_examples

reconnaissance faciale d'Apple qui peine à identifier les visages asiatiques. Les exemples les plus divers de défauts présentés par des systèmes d'IA sont légions. Leur impact est certes variable, mais ils tendent à démontrer que la faillibilité de la machine est une composante inhérente de celle-ci.

4. Le danger démagogique

Pour Serge Tisseron et Frédéric Alexandre, le danger principal que l'intelligence artificielle pourrait faire courir à terme à l'Humanité n'est pas tant celui dénoncé par les auteurs des multiples lettres ouvertes dont nous avons parlé. Il se situe ailleurs et se révèle plus intangible. Ses conséquences dans le domaine de la décision politico-militaire, par contre, pourraient être redoutables. Pour comprendre les propos de Tisseron et Alexandre, il importe de nous attarder quelque peu sur la définition que nous donnons de l'intelligence. Il existe, en réalité, deux facettes principales à l'intelligence : l'intelligence formelle et l'intelligence émotionnelle. L'intelligence formelle se rapporte aux traitements mathématiques, combinatoires, logiques et statistiques. Cette intelligence rassemble les processus rationnels tels que ceux qui conduisent aux déductions, aux inférences logiques. Chez l'homme, l'intelligence formelle mobilise le cortex préfrontal. À côté de l'intelligence formelle, l'intelligence émotionnelle convoque le système limbique. C'est la partie de l'intelligence qui pousse l'homme à se protéger des risques, à développer des stratégies de survie, à se reproduire, à se nourrir, à entretenir des rapports sociaux, à nouer des alliances, etc. En termes plus simples, il s'agit de l'intelligence qui guide l'homme dans son environnement en tentant de lui permettre de surmonter les contraintes qui y sont liées. Ce faisant, cette intelligence émotionnelle va conduire l'homme à attribuer des valeurs aux choses. En fonction de ces valeurs, le système limbique va influer sur les meilleures décisions à adopter afin de diminuer la part de risques et d'optimiser les gains. Contrairement à une vision cartésienne « absolutiste », il n'existe pas de séparation entre le corps et l'esprit. L'intelligence formelle (celle du raisonnement, des mathématiques, de la logique) entretient une liaison permanente avec le système limbique (creuset de l'intelligence émotionnelle et de l'attribution des valeurs) qui oriente l'individu dans le monde réel.

Quel que soit le degré de développement de l'intelligence artificielle dans l'état actuel des technologies, celle-ci est avant tout une simulation d'intelligence. Certes, une IA peut aujourd'hui permettre la réalisation d'opérations hautement complexes que l'homme ne pourrait être en mesure de réaliser, et encore moins à la vitesse à laquelle l'IA travaille. Toutefois, il ne s'agit là encore et toujours que d'une simulation d'intelligence. Pourtant, une simulation peut générer une illusion, celle de l'intelligence précisément, et provoquer chez l'homme des réflexes inattendus. Ainsi, la chercheuse Julie Carpenter de l'Université de Washington s'est-elle penchée sur l'empathie naissante parmi les soldats de l'U.S. Army à l'égard des robots intégrés dans leurs unités. Elle s'est plus spécifiquement intéressée au cas du personnel militaire des équipes de déminage qui emploient des systèmes robotiques pour le désamorçage d'explosifs sur les théâtres d'opération. Il s'agissait d'étudier le rapport qu'entretiennent les soldats avec leurs robots et de déterminer si la nature et l'importance de ce rapport était susceptible d'avoir une incidence sur la qualité du travail des équipes et sur la sécurité du personnel. Les conclusions auxquelles a abouti la chercheuse sont étonnantes à plus d'un titre. Tout d'abord, elle a pu constater que la nature des relations qu'entretiennent les membres du personnel militaire avec les robots de leurs unités évolue en fonction du degré d'avancement technologique des robots. Bien que les soldats avec lesquels Julie Carpenter s'est entretenue ont affirmé ne pas avoir un attachement réel vis-à-vis de leurs robots, ils ont également avoué être tristes, parfois même ressentir de la frustration ou de la colère lors de la perte de certains exemplaires lors d'opérations de déminage.

Le principal danger, aux yeux de Serge Tisseron et Laurent Alexandre, ne réside donc pas – contrairement à ce que voudraient nous faire croire les partisans de l'IApocalypse – dans l'émergence d'une intelligence éminemment supérieure à celle des hommes ou, pour le dire

autrement, le véritable risque que pourrait courir l'Humanité ne découle pas des performances combinatoires et mathématiques des intelligences artificielles existantes et futures. D'ailleurs, dans nombre de domaines, ce sont des systèmes d'IA aux capacités logiques de très loin supérieures à celles de l'homme qui sont employés. Quand Elon Musk, Stephen Hawking ou Bill Joy insistent sur les périls que pourraient générer les systèmes d'IA de demain, ils concentrent en réalité leur propos sur les aptitudes mathématiques des systèmes d'IA. Or, selon Tisseron et Alexandre, il n'est nul besoin de s'alarmer de cet aspect de l'IA.

Où donc se situe le véritable risque de l'IA ? Chez l'homme, affirment Tisseron et Alexandre. Comme l'étude sur le rapport entre le soldat et le robot a pu l'illustrer, la valeur d'une machine n'est qu'une valeur « prêtée », c'est-à-dire qu'elle découle avant tout du regard que porte l'homme sur la machine. Il nous faut ici revenir aux notions d'*explicabilité* et de *compréhensibilité* de l'IA (cf. plus haut). L'homme aura d'autant plus tendance à « valoriser » un système d'IA qu'il sera de moins en moins en mesure de comprendre les processus d'inférence profonds qui se situent à la base de son fonctionnement et des résultats qu'elle produira. En des termes plus simples, plus le fonctionnement d'un système d'IA lui paraîtra complexe, plus l'homme sera tenté de considérer ce système comme lui étant supérieur et plus il lui accordera de « valeur ». Ce biais cognitif ira en s'amplifiant avec les systèmes d'IA basés sur l'apprentissage machine. En effet, le niveau de performance futur des systèmes d'IA sera la résultante directe de la qualité de l'apprentissage, qualité qui dépendra elle-même de la qualité et de la quantité des masses de données sur lesquelles les systèmes d'IA se seront entraînés. À l'avenir, nous aurons donc affaire, dans une multiplicité de champs d'action et de secteurs d'activité, à des systèmes d'IA qui auront une connaissance particulièrement fine de nos attentes, qu'il s'agisse d'attentes d'ordre individuel ou de niveau collectif. Le danger réel est de voir apparaître des systèmes d'IA « démagogues » dont les conseils ou les décisions se conformeraient bien plus aux souhaits de l'individu ou de la collectivité qui les requiert qu'aux données réelles du problème que ces système d'IA seraient supposés résoudre[111]. Ce type de risque est, nous en convenons, difficile à concevoir de manière concrète. Pourtant, il pourrait donner lieu à des problématiques très réelles dans le domaine militaire, notamment en matière de commandement. Quelle serait par exemple, demain, la valeur d'une armée si les systèmes d'IA qui appuieront et aviseront les instances de commandement et de contrôle délivraient des conseils fondés davantage sur la gratification ou la satisfaction de leur « client » que sur les données issues de l'espace de bataille et du rapport de forces entre les protagonistes ? Serait-il concevable, dans un futur plus ou moins proche, d'imaginer des chefs d'État et de gouvernement majoritairement conseillés par des systèmes d'IA qui auraient chacun pour fonction de maximiser les gains stratégiques des seules élites qu'ils conseillent ? Ces mêmes élites toléreraient-elles d'être appuyées par des systèmes d'IA qui, plutôt que de rechercher à tout prix l'accumulation des bénéfices stratégiques, économiques et politiques des dirigeants qu'ils conseillent, délivreraient des solutions se limitant à la recherche d'un optimum calculé sur la base de la place et du rang de chaque État dans le système international ?

Comme l'évoquent Tisseron et Alexandre, le « problème important n'est pas l'existence de machines surpuissantes, mais de savoir qui leur fixe des objectifs, des valeurs et des limites. Et jusqu'à présent, ce rôle est toujours dévolu à l'être humain.[112] » On attribue trop souvent à la machine des

[111] Dans le film *Her*, un héros entretient une relation amicale, puis amoureuse, avec une intelligence artificielle nommée *Samantha*. Cette IA parvient à se construire une connaissance parfaite du héros grâce à un accès sans limite à toutes les données qui le concernent sur l'internet. Par ailleurs, son rôle de « relais » et de confidente lui permet d'être en parfaite osmose avec les ressentis du héros. Grâce à cette faculté, elle finit par lui imposer ses propres choix. *Cf.* Frédéric ALEXANDRE, Serge TISSERON, « Où sont les vrais dangers de l'intelligence artificielle ? », *Pour la Science*, Dossier numéro 87, avril-juin 2015, p. 104.

[112] *Ibid.*, p. 106.

caractéristiques qui ne sont aucunement liées à l'expérience de la machine mais bien à la façon dont l'homme a programmé la machine pour que celle-ci prenne en considération certaines variables plutôt que d'autres pour l'élaboration de sa représentation du monde. La véritable question qui se posera à l'horizon des dix prochaines années sera de déterminer la philosophie et les valeurs sur lesquelles les systèmes d'IA seront bâtis. Il appartient à l'homme de ne pas créer une intelligence artificielle qui ne soit là que pour sa propre gratification et puisse englober l'intérêt général dans ses paramètres.

5. Le « transmachinisme » : faire évoluer les machines indépendamment de l'homme

Notre relation à la machine, comme tendent à le montrer les débats dont nous avons essayé de faire émerger la substance, s'est largement construite selon une dialectique d'opposition quelque peu paradoxale. La volonté de développer des machines intelligentes sur le modèle de l'homme (émanant majoritairement des scientifiques et des ingénieurs, avec l'espoir si peu secret de dépasser ce dernier) s'est heurtée à de nombreuses reprises à la réalité sociale d'individus-citoyens voyant dans les phases successives de l'évolution de l'IA et de la robotique des menaces pour la place et la centralité de l'homme au sein des organisations. De la même façon, le projet soutenu par le courant transhumaniste de modifier la physiologie de l'homme afin d'en augmenter les capacités au point de le faire évoluer au niveau de la machine fait aujourd'hui polémique. Si les débuts du transhumanisme pouvaient laisser sourire les observateurs scientifiques et certains philosophes, l'énoncé des principes fondateurs du courant à l'aune des progrès technologiques des vingt dernières années peut inquiéter. Quels sont-ils ? Il y a tout d'abord la conviction que l'Humanité est à la veille de transformations majeures du fait de la technologie. La possibilité d'améliorer, de rajeunir, de réparer le corps humain serait bientôt à portée de main, grâce notamment aux technologies NBIC. Plus encore, l'homme disposera demain des moyens qui lui permettront d'accroître son intelligence, de réduire l'emprise de son inconscient sur son équilibre psychologique, d'abolir toute notion de souffrance. C'est donc le dépassement de la biologie humaine qui deviendrait une sorte de droit fondamental pour l'homme.

Plutôt que d'envisager le rapport entre l'homme et la machine selon une logique de compétition visant à savoir qui de l'homme ou de la machine dominera l'autre dans un seul et même organisme (biologique ou artificiel), Jean Rohmer énonce une autre voie dans laquelle l'homme et la machine, tout en étant dans un environnement de coexistence, pourraient évoluer séparément mais en symbiose. La posture défendue par Jean Rohmer est également appelée « transmachinisme ». L'idée fondatrice de l'auteur est « d'imaginer une évolution des machines et de l'industrie en général non pas pour dépasser ou transformer l'homme, mais pour permettre aux machines de mieux faire leur travail de machines[113] ». L'idée défendue par le transmachinisme est de pousser les capacités de travail autonome des machines au maximum de leur utilité. Il ne s'agit pas toutefois d'envisager une machine s'approchant de ou dépassant l'intelligence humaine, pas plus qu'il n'est question de réfléchir à la façon dont l'homme pourrait se transformer pour ressembler à une machine ou en devenir une. Dans la vision transmachiniste, l'homme et la machine restent et demeurent deux entités – ou plutôt groupes d'entités – distinctes. Cette approche s'inspire, au demeurant, de certains exemples de la réalité. On citera, notamment, les camions Volvo de la mine de Kristineberg en Suède qui se relaient de manière autonome afin de permettre l'évacuation du minerai et des déchets[114]. Ou encore, en Norvège cette fois-ci, le projet de navires autonomes développé par les sociétés Kongsberg et Rolls Royce. Le transmachinisme envisage le développement de « bulles productrices autonomes »

[113] Jean ROHMER, « Le transmachinisme : et si les machines évoluaient indépendamment de l'homme ? », *The Conversation*, *cf*. https://theconversation.com/le-transmachinisme-et-si-les-machines-evoluaient-independamment-de-lhomme-138367
[114] https://www.volvogroup.fr/fr-fr/news/2016/sep/news-2297091.html.

indépendantes de l'homme et au sein desquelles les machines exercent leurs fonctions de machines. Un monde transmachiniste s'articulerait autour de divers sous-ensembles de production installés, supervisés, entretenus et recyclés par des machines. Celles-ci seraient également capables de produire leurs propres sources d'énergie et développeraient leurs propres infrastructures de gestion des productions qu'elles auraient pour mission de réaliser. Les bulles productrices autonomes opéreraient dans une forme d'autogestion continue[115].

Jean Rohmer évoque, au-delà de la présentation de l'approche transmachiniste, les futurs possibles d'une société dans laquelle les machines évolueraient aux côtés des hommes sans pour autant que ces deux mondes s'interpénètrent. Une première évolution vers laquelle tendraient les machines reposerait sur l'indépendance énergétique et financière visant l'optimisation des moyens et capacités de production. Ensuite, et toujours dans la perspective d'une optimisation de la production au service de l'homme, les machines acquerraient une compréhension du langage humain. Plus encore, elles se rendraient capables d'envisager la totalité des connaissances humaines dans un ensemble cohérent et formalisé depuis les multitudes de données disponibles à travers l'internet. Aucun processus de fabrication, aucun modèle de conception de machines, aucune théorie scientifique, aucune compétence d'ingénierie n'échapperait aux machines. Enfin, fortes de l'ensemble des connaissances acquises grâce à l'homme, les *transmachines* élaboreraient leurs propres connaissances et théories.

Un monde transmachiniste verrait l'homme et la machine évoluer côte à côte tout en s'ignorant mutuellement. On peut supposer que l'évolution indépendante des machines dans un monde où nulle confusion entre l'homme et la machine n'existerait pourrait aboutir à l'émergence de solutions et de modes de raisonnement bien différents de ceux découlant de notre génie mécanique et civil. On sait, par exemple, à quel point les programmes informatiques et algorithmes conçus par l'homme regorgent d'erreurs, de désordres et de biais.

6. Conclusion partielle

Le débat de ces cinq dernières années au sujet des systèmes autonomes et de l'IA a considérablement pâti des diverses mises en garde, alertes et visions catastrophistes qui ont pourtant émané de technologues réputés ou de scientifiques confirmés et de renom. De telles postures ne présentent que peu d'originalité et sont parfois empreintes des symboles véhiculés par l'imaginaire de la science-fiction. Elles dépendent dans une large mesure des a priori qui peuvent, selon les époques, affecter le rapport qu'entretient une société avec la technologie. Qu'ils soient scientifiques, ingénieurs ou industriels, les experts qui expriment une vision de l'avenir restent, parfois inconsciemment, tributaires des préjugés du temps auxquels ils appartiennent. Cette mise en garde ne signifie nullement que les efforts d'anticipation et autres alertes formulées soient dénués du moindre intérêt. Faute de posséder une valeur stratégique exploitable, de telles mises en garde nous informent de l'état d'esprit qui peut exister dans une société donnée à une époque donnée. On observera que les réticences formulées à l'égard du potentiel de l'IA émanent principalement, pour ne pas dire exclusivement, du monde occidental. L'Asie, notamment la Chine, ne semble pas s'inscrire dans l'approche apocalyptique véhiculée par l'Occident. C'est là un point qui n'est nullement un détail et dont les implications géopolitiques peuvent s'avérer déterminantes.

Les multiples évocations de futurs qui déchantent du fait des progrès de l'IA traduisent surtout le malaise d'une communauté scientifique par rapport à la nature et à l'importance des évolutions réelles

[115] Les tenants du transmachinisme affirment que de nombreux sites de production et d'opérations fonctionnent déjà selon leurs critères. Ainsi, en Chine par exemple, les grands ports chargent et déchargent des conteneurs sur des quais vidés de toute présence humaine. De la même façon, la production et la fabrication de puces électroniques sont aujourd'hui des processus quasi entièrement automatisés.

des technologies du numérique. Elles reflètent également l'état schizophrénique d'une communauté de développeurs et d'utilisateurs qui, tout en se montrant de plus en plus dépendante des technologies de l'IA et de ses dérivés au quotidien, semble aussi en proie à des inquiétudes existentielles sur l'avenir des rapports entre l'homme et la machine.

V. Géopolitiques de l'intelligence artificielle

Dans la conclusion de leur article consacré à l'étude du concept de l'hyper-guerre, Allen et Husain pointaient la nécessité pour les États-Unis de mobiliser les moyens nécessaires aux investissements stratégiques qui leur permettront de mener demain des confrontations avec leurs adversaires futurs. Cette interpellation s'inscrit, en vérité, dans un concert de déclarations tendant à placer dans la maîtrise de l'IA le devenir des équilibres militaires, de la sécurité et de la paix. Ainsi, en septembre 2017, le président russe Vladimir Poutine exprimait-il sa vision géopolitique d'un monde à l'ère de l'IA : « celui qui deviendra leader en ce domaine sera le maître du monde ». Quelques jours plus tard, Elon Musk, fondateur de SpaceX et Tesla, n'hésitait pas à s'inquiéter des répercussions sécuritaires de l'IA en exprimant sa crainte que la lutte entre nations pour la supériorité en matière d'IA ne soit le déclencheur de la troisième guerre mondiale.

L'intérêt croissant que suscite l'IA dans les forces armées du monde entier nous incite à nous interroger sur les perspectives et les risques d'une possible prolifération des technologies dans ce secteur. Si la littérature stratégique évoque l'intelligence artificielle comme l'un des domaines applicatifs qui pourraient modifier les équilibres militaires du siècle, force est de constater que les niveaux d'investissement des États peuvent varier considérablement. Il est donc utile de fournir un bref aperçu des principaux pays leaders de cette technologie dans le secteur de la défense – au sein et à l'extérieur de l'Alliance –, ainsi que des initiatives engagées par l'OTAN.

Il va sans dire qu'une compétition à l'échelle mondiale existe bel et bien. Elle entremêle des acteurs publics (les États) et des acteurs privés (des entreprises innovantes, des start-ups d'un genre nouveau, parfois soutenues par les États). Encore faut-il comprendre les caractéristiques de cette compétition. Les applications de l'intelligence artificielle portent sur des marchés extrêmement diversifiés et en pleine croissance à l'échelle internationale. De nombreux acteurs du numérique, comme nous avons pu l'observer, ont développé depuis plusieurs années des stratégies d'investissement, mais aussi des politiques de soutien ou d'acquisition de petites et moyennes entreprises innovantes dans des niches spécifiques. Combiné aux capacités de récolte et de stockage de *big data* de ces grands groupes, l'apport de ces start-ups dans les secteurs du *data mining* ou du *deep learning* débouchera sur des ensembles d'activités extrêmement polyvalents dont les produits pourront concerner des applications en matière de défense. Il est donc particulièrement difficile aujourd'hui de recenser avec exactitude les acteurs de cette compétition globale qui influenceront les équilibres militaires de demain.

Si l'attention d'un large public tend à se concentrer sur les relations (jugées à tort évidentes) entre l'IA et la robotique, il importe d'élargir considérablement l'ampleur du spectre applicatif dans lequel l'IA jouera un rôle dans les domaines de la sécurité et de la défense. Ainsi, certaines armées travaillent à la mise au point de systèmes d'aide à la décision, tant au niveau stratégique que sur le plan tactique. Confrontées au stress et au rythme particulièrement élevé des opérations militaires modernes, les armées à haute capacité technologique choisissent de s'aider de machines dont la fonction est de produire des analyses situationnelles plus rationnelles et capables d'intégrer un grand nombre de variables que les opérateurs humains ne seraient pas en mesure d'intégrer en une seule et même représentation d'un espace de bataille. La rapidité d'analyse, de prise de décision et de mise en œuvre des décisions constitue un atout essentiel dans la compétition que se livrent les puissances d'aujourd'hui. L'apport de tels systèmes d'aide à la décision – frappant bien moins l'imaginaire – est fondamental pour les États désireux d'atteindre une supériorité dans la décision militaire.

Les démarches engagées par de nombreux États à travers la planète en attestent. Dans la dernière décennie, de nombreux États et groupes d'États ont amorcé d'importantes réflexions concernant l'impact de l'IA sur les systèmes de défense[116]. En 2017, le Stockholm International Peace Research Institute (SIPRI) avait comptabilisé l'existence de pas moins de 49 systèmes d'armes autonomes incorporant de l'IA embarquée capables d'engager une cible sans recours à l'assistance humaine. Et ce ne sont là que les prémices d'une course technologique aux armements bien plus vaste[117]. La question est désormais posée par les experts : l'IA se confirmera-t-elle comme la technologie au cœur d'une rivalité globale nouvelle entre les superpuissances du XXIe siècle que seront les États-Unis et la Chine, tout comme le fut la technologie nucléaire et thermonucléaire dans le rapport de force de la guerre froide entre Washington et Moscou au cours du XXe siècle[118] ?

Avant tout, il convient de rappeler la place qu'occupera l'IA demain au sein des États pour leurs besoins militaires. L'IA constituera moins une arme qu'un multiplicateur de force, comme le furent en leurs temps respectifs l'électricité, la radio, le radar ou encore, plus récemment, les capacités de types C4ISR[119]. En tant que nouvelle catégorie technologique, les systèmes d'IA pourraient redéfinir et transformer l'actuel équilibre des forces à l'échelle du monde entre les puissances industrielles avancées. D'une manière générale, les armées qui intégreront au sein de leurs dispositifs militaires des moyens dotés d'IA pour l'assistance à la prise de décision disposeront d'un avantage comparatif certain par rapport aux organisations militaires qui dépendront quant à elles de la seule décision humaine. Surtout, ce sont les organisations de défense assistées par l'IA qui détermineront demain le rythme stratégique, opérationnel et tactique des opérations militaires[120].

On prendra soin de distinguer les impacts de l'IA dans le champ politico-militaire selon les niveaux considérés. Sur le plan stratégique, ce sont les mécanismes de prise de décision qui seront principalement affectés. Les systèmes de commandement et de contrôle (C2) assistés par IA s'affranchiront très certainement d'un nombre important d'interférences liées à la dépendance des procédures et des planifications à la variable humaine. Nombreux sont ceux qui voient l'IA comme une solution au brouillard de la guerre qui affecte la conduite des forces en temps de crise : émotion liée à l'engagement de forces ou à l'évaluation des coûts humains, biais résultant d'une réflexion de groupe, impact de la hiérarchie sur la rationalité de la prise de décisions, etc. Comme nous aurons l'occasion de l'observer dans la suite de cette étude, de nombreux programmes aux États-Unis sont actuellement engagés pour réduire le facteur humain au sein des procédures décisionnelles stratégiques et militaires. Certes, il ne s'agit pas d'arracher complètement l'homme de la chaîne décisionnelle, mais d'atténuer autant que faire se peut l'indétermination de la variable humaine sur les processus d'engagement des forces. Plus encore, la recherche militaire dans le domaine de l'IA vise à permettre une anticipation aussi en amont que possible des signaux faibles. Cette capacité de détection, d'analyse et de traitement des informations issues de diverses sources de renseignement s'inscrit dans une double démarche de prévention et de préemption stratégique.

---

[116] Bob WORK, (14 décembre 2015), « Deputy Secretary of Defense Bob Work's Speech at the CNAS Defense Forum, retrieved from http://www.defense.gov/News/Speeches/Speech-View/Article/634214/cnas-defense-forum

[117] Michael C. HOROWITZ, (15 mai 2018), « Artificial Intelligence, International Competition, and the Balance of Power », Texas National Security Review, retrieved from https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/

[118] Gregory ALLEN & Elsa KANIA, (8 septembre 2017), « China is using America's own plan to dominate the future of artificial intelligence », Foreign Policy, retrieved from https://foreignpolicy.com/2017/09/08/china-is-using-americas-own-plan-to-dominate-the-future-of-artificial-intelligence/

[119] Computerized Command, Control, Communications, Intelligence, Surveillance and Reconnaissance.

[120] James JOHNSON, « Artificial Intelligence & Future Warfare: Implications for International Security », *Defense & Security Analysis*, Vol. 35, No. 2, 2019, p. 150.

1. Vers une domination sino-américaine ?

Au-delà ou en deçà (selon le point de vue que nous adoptons) de la crainte d'une lutte violente entre les nations pour la supériorité dans les technologies de l'IA, l'émergence future d'un *condominium* – ne serait-ce que provisoire – de la Chine et des États-Unis dans le domaine de l'IA constitue une hypothèse d'évolution probable de la géopolitique mondiale. Les empires numériques américain et chinois exerceront très certainement une régulation et une domination dans les années à venir. Leur maîtrise de l'IA constituera à n'en pas douter l'atout majeur de leur emprise sur le devenir des sociétés. Certains auteurs ne manquent d'ailleurs pas de considérer que l'IA se confirmera, dans les décennies futures, comme le levier de nouveaux empires. Certes, il s'agira là d'empires d'un genre quelque peu nouveau, dans la mesure où leur structure s'articulera autour de sociétés multinationales échappant aux normes établies par les nations, alors même que ces dernières auront entre-temps financé de larges pans des activités de ces multinationales avant qu'elles ne prennent leur essor et leur indépendance.

Si les premiers développements de l'internet et des technologies numériques avaient pu laisser entrevoir l'espoir d'une décentralisation toujours plus grande sur le plan économique et politique, la croissance de l'IA et des technologies qui lui sont associées devraient en réalité conduire le monde vers une centralité des structures économiques et sociales au point que seuls quelques acteurs privés détiendront demain les clés des dispositifs de régulation politique. L'héritage historique des États-nations associé à la puissance des groupes privés dominant les technologies de l'IA tendrait donc à aboutir à un système international hybride où l'on pourrait retrouver, avec des contours quelque peu différents, des logiques de blocs et de non-alignement. La masse de données et les capacités de traitement qu'implique le développement de l'IA enclenchera, presque nécessairement, une double dynamique monopolistique et oligopolistique : monopolistique dans le cadre d'un espace politico-économique donné et oligopolistique à l'échelle mondiale (plusieurs entreprises innovantes parvenues à dominer les technologies de l'IA se partageant le marché mondial). On ajoutera encore que cette domination mondiale sera facilitée par la spécificité dans la transposition des algorithmes entre machines. Les algorithmes du *machine learning* utilisés dans de nombreuses applications sont transférables à d'autres selon un processus appelé « apprentissage par transfert ». Cette méthode d'extension des lignes de programme composant les machines permet une extension rapide des dispositifs apprenants. La domination monopolistique/oligopolistique que nous évoquons ici se traduit par ailleurs par un rapprochement toujours plus grand entre les fabricants de matériel (*hardwares*) et de programmes (*softwares*). Si le développement de l'informatique durant les deux dernières décennies a principalement reposé sur une division du travail entre, d'une part, les concepteurs de logiciels et, d'autre part, les fabricants de matériels permettant le fonctionnement de ces logiciels, il semble que l'économie numérique à l'ère de l'IA pousse vers une convergence entre *hardware* et *software*. En d'autres termes, les concepteurs de programmes optent pour le développement de leurs propres solutions matérielles dans la recherche d'une adéquation toujours plus précise entre le logiciel et son support. Cette tendance génère, à son tour, une centralisation encore plus poussée des acteurs numériques à l'ère de l'IA.

Il importe d'insister tout particulièrement sur la place qu'occuperont dans ce système international les groupes privés dominant les technologies de l'IA, d'une part, et sur le relationnel dissymétrique que ces derniers entretiendront avec les élites étatiques, d'autre part. La question de la régulation de ces nouveaux acteurs privés est désormais posée et ne semble pourtant susciter aucune réponse satisfaisante. Il nous faut, en effet, réaliser que ces groupes d'acteurs emploient des équipes d'ingénieurs chargées de la constitution de grands ensembles de données produites par tout un chacun : consommateurs, vendeurs, travailleurs, usagers divers, citoyens, etc. Ces groupes élaborent, évaluent et paramètrent des algorithmes, interprètent les résultats obtenus et définissent la façon dont ces algorithmes seront utilisés au sein de la société. Si ce sont, en vérité, les usagers

(économiques, politiques, culturels, etc.) qui alimentent les systèmes d'IA en données, ce sont bien les groupes technologiques dominants qui exploitent les « temps de cerveau disponibles » que nous leur offrons de la sorte[121].

La réalité de l'existence d'une compétition internationale dans le domaine des technologies de l'IA est un fait que confirme l'agitation des États en matière de politique scientifique et technologique durant la décennie écoulée. Rares sont les États ou entités étatiques à ne pas avoir publié leur propre stratégie ou feuille de route en matière d'IA : la France, le Canada, la Chine, le Danemark, l'Union européenne, la Finlande, l'Inde, l'Italie, le Japon, le Mexique, les régions scandinave et balte, Singapour, la Corée du Sud, la Suède, Taïwan, les Émirats arabes unis ou encore le Royaume-Uni : tous ont, en quelques mois, rédigé diverses stratégies tantôt destinées à promouvoir l'utilisation et le développement de l'IA, tantôt conçues en vue de mettre sur pied les infrastructures devant permettre l'évolution des politiques d'éducation et de recherche jugées nécessaires afin de s'aligner sur la tête de peloton des puissances scientifiques maîtrisant les technologies de l'IA. Certes, les politiques envisagées peuvent différer selon les États considérés. Pour certains, l'ambition est de devenir de acteurs globaux dans le domaine et de maîtriser l'ensemble des technologies, capacités et filières de l'IA : c'est le cas, notamment, des États-Unis, de la Chine et, plus tardivement, de l'Union européenne. Pour d'autres acteurs internationaux, ce qui est visé est la maîtrise de niche. En d'autres termes, il s'agit de parvenir à posséder des compétences clés qui puissent répondre aux intérêts d'acteurs spécifiques. C'est notamment le cas de l'Inde, dont l'ambition est de devenir un « garage de l'IA » en développant des applications spécialement consacrées aux pays en développement. La Pologne, quant à elle, cherche à accroître ses compétences dans le domaine sensible de la cybersécurité et, plus généralement, de l'IA à des fins militaires.

Bien qu'il soit très difficile d'entrevoir aujourd'hui les transformations qu'induiront les technologies de l'IA sur les structures de forces des armées des nations les plus industrialisées, plusieurs remarques peuvent être néanmoins faites sur la façon dont ces nations – en premier lieu desquelles la Chine et les États-Unis – se sont engagées sur la voie d'une intégration toujours plus poussée de l'IA au sein de leurs édifices militaires. Comme toute rupture technologique majeure, l'IA connaîtra une intégration par paliers. Les précédentes révolutions technologiques militaires le montrent : les mutations qu'elles ont engendrées au sein des organes de défense ont connu de nombreuses itérations. Le processus d'adaptation des doctrines aux ruptures technologiques est un processus lent et lourd, qui se heurte souvent au conservatisme « naturel » des organisations. À ce phénomène d'inertie s'ajoutent bien d'autres facteurs permettant d'expliquer la lenteur des mécanismes adaptatifs des forces armées. Les pesanteurs idéologiques, les particularités des cultures stratégiques forgées au gré de multiples expériences historiques sont autant de mécanismes qui sont susceptibles d'affecter la traduction optimale d'une technologie nouvelle au sein d'un système de forces.

Ainsi, les premières approches de la Chine dans le domaine de l'IA ont-elles été fortement marquées par l'évaluation des initiatives militaires américaines, en particulier celles associées à la troisième stratégie de compensation (*Third Offset Strategy*). Ce n'est que récemment que la Chine a développé une vision propre et plus mature de l'intégration de l'IA au sein de son dispositif militaire. L'Armée populaire de libération (APL) possède, en effet, une organisation spécifique, associée à des pratiques et des cultures de services particulières. La compétition que se livreront sans nul doute la Chine et les États-Unis en ce qui a trait à l'implantation des systèmes d'IA au sein de leurs dispositifs respectifs de défense ne résidera pas en une concurrence entre solutions similaires. Chaque entité a ses particularités, ses cultures de travail, ses procédures et ses normes hiérarchiques. Qui plus est, les biais

[121] Nicolas MIAILHE, « Géopolitique de l'Intelligence artificielle : le retour des empires ? », *Politique étrangère*, volume 3, 2018, pp. 105-117.

inhérents à toutes ces spécificités pourraient, en cas de crise ou de tension, avoir des répercussions sur la construction même des décisions qui seraient élaborées sur la base des données analysées et traitées par des systèmes d'IA.

2. Les États-Unis

C'est très rapidement que les États-Unis ont évoqué (sans encore employer l'expression aujourd'hui consacrée, toutefois) le rôle futur de l'*intelligence artificielle* dans le domaine militaire. En 2003, un rapport du Congrès sur la *Transformation* évoquait déjà les perspectives d'un recours au *data mining*. L'intelligence artificielle était alors seulement entrevue comme une piste technologique nécessitant du temps avant d'être maîtrisée à des fins opérationnelles militaires. D'une certaine façon, le recours à l'intelligence artificielle dans les armées n'est que le dernier avatar d'un processus de refonte de l'outil doctrinal entamé dans le cadre de la guerre froide sous la forme d'une *Offset Strategy* (stratégie de compensation). De telles stratégies de compensation émergent de manière cyclique sur la base de considérations relevant principalement d'un prisme psychologique de la part des responsables politico-militaires américains. La première stratégie de compensation des États-Unis, dans les années 1950, était née du constat d'une infériorité américaine dans le domaine des forces conventionnelles, principalement sur le théâtre européen. Face à cette situation, la stratégie des États-Unis a consisté à renforcer ses moyens balistiques nucléaires tactiques. Une seconde stratégie de compensation est adoptée à la fin des années 1970 tout à la fois sous la pression de l'expansion soviétique dans les pays du tiers-monde et de la nécessité de se relever de la guerre du Vietnam. Cette seconde *Offset Strategy* s'appuiera sur les technologies des ordinateurs et des réseaux. Il s'agit, en d'autres termes, de concevoir la maîtrise de l'information comme une arme devant assurer la supériorité des forces armées américaines. Jointes aux technologies de précision qui équiperont les missiles, cette révolution informationnelle des forces armées s'avèrera décisive lors de l'opération coalisée *Tempête du Désert* en 1991 contre les forces armées irakiennes : la *Revolution in Military Affairs* (RMA) était née et allait assurer une domination militaire américaine durant près de deux décennies.

Au tournant de l'année 2010, une contestation progressive de la supériorité technologique américaine voit le jour. Elle résulte de la conjugaison de plusieurs phénomènes. Le premier est la fatigue des forces armées des États-Unis du fait d'une sur-expansion stratégique résultant des campagnes militaires d'Afghanistan et d'Irak. Les opérations *Enduring Freedom* (2001) et *Iraqi Freedom* (2003), si nous les envisageons dans leur seule phase opérationnelle, furent des succès militaires indéniables. Elles furent, toutefois, suivies d'occupations de territoires qui se révélèrent extraordinairement problématiques et pour lesquelles les États-Unis échouèrent à adopter des stratégies claires en matière de sortie de crise. L'épuisement des forces américaines sur ces deux terrains permirent à des États comme la Russie et la Chine de tirer les leçons pour, d'une part, définir le type de force et de stratégie à développer contre les États-Unis et, d'autre part, éviter les erreurs stratégiques commises par Washington sur des théâtres distants.

La compétition géopolitique engagée entre les États-Unis et le « nouvel entrant » que représente la Chine sur le plan des relations internationales s'est rapidement traduite, en l'absence de confrontation militaire, par une concurrence ardue entre ces deux nations sur le plan de la maîtrise des technologies avancées dans le secteur numérique. Il est essentiel de comprendre qu'à l'intérieur même de ces deux puissances, de nouveaux acteurs ont émergé dans le domaine numérique qui rivalisent actuellement avec les groupes industriels « historiques ». Aux États-Unis, les GAFAMITIS (Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce) et les NATU (Netflix, Airbnb, Tesla et Uber) bataillent contre les géants chinois que sont les BATX (Baidu, Alibaba, Tencent et Xiaomi[122]). Tant aux

[122] On pourrait ajouter la société Huawei.

États-Unis qu'en Chine, ces nouveaux entrepreneurs du numérique ont acquis une position dominante dans chacun de leur domaine de développement. Ils ont imposé des normes nouvelles grâce aux contraintes liées à l'interopérabilité. Leur force réside, plus exactement, dans leur clientèle dont ils exploitent les données d'utilisation dans tous les domaines de la vie quotidienne. Les technologies déployées par ces nouvelles entreprises non traditionnelles servent à la récolte de données, à leur traitement et à leur analyse. Il existe une véritable valorisation des informations issues du cadre dans lequel les clients évoluent. C'est là l'atout principal des nouveaux entrepreneurs du numérique ; c'est là que se situe la source de l'explosion de l'IA depuis 2010. Si la Chine dispose d'une avance certaine en matière de récolte de données grâce à l'importance de sa population et du nombre d'utilisateurs des technologies numériques sur son sol (la Chine ambitionne de posséder 30 % des données mondiales d'ici 2030), elle souffre d'un manque d'ingénieurs spécialisés dans l'IA. Avec ses 39.000 chercheurs/ingénieurs, la Chine ne possède que la moitié du bassin d'emplois américains spécialisés dans l'IA, soit 78.000 chercheurs. Il faut encore ajouter que la Chine reste dépendante des États-Unis pour le développement des processeurs et des puces et, plus exactement, des processeurs graphiques GPU (*Graphical Processor Units*) indispensables pour le *machine learning*.

En tant que numéro un mondial de l'IA, et afin de se maintenir comme leader, les États-Unis ont activement cherché à intégrer la technologie de l'IA dans leurs capacités militaires. Le développement d'une IA adaptée au secteur de la défense représente toujours une bonne partie des efforts qui avaient été amorcés dans le cadre de la *Third Offset Strategy* de l'administration Obama, dont l'objectif était de préserver l'avance militaire du pays. En 2018, le DoD a dévoilé une stratégie dans le domaine de l'IA. Elle a été accompagnée de l'*American AI Initiative* sous l'administration Trump.

Le DoD a adopté une politique d'investissements massifs dans le secteur de l'IA. Ainsi, entre 2013 et 2017, ce ne sont pas moins de 1,76 milliard de dollars américains qui furent consacrés à trois postes pertinents du budget de la défense (apprentissage et renseignement ; informatique de pointe ; systèmes à base d'IA)[123]. En 2018, la DARPA a annoncé un supplément budgétaire de quelque 2 milliards de dollars pour la période 2018-2023. En 2016, la *Defense Innovation Unit* (DIU) voit le jour. Son rôle : faciliter l'intégration de la technologie du secteur privé dans le domaine de la défense. L'IA est désormais un domaine clé des activités de cette unité. En 2018, le Department of Defense établit le *Joint Artificial Intelligence Center* (JAIC) sur la base des recommandations produites quelques années plus tôt à propos de la nécessité de centraliser les activités liées à l'intégration de l'IA au sein des forces armées. Le JAIC dispose d'un budget en progression exponentielle : de 75 millions de dollars pour ses débuts, l'investissement a été porté à 1,75 milliard de dollars sur 5 ans pour superviser et coordonner les actions du ministère en matière d'IA.

a) La spécificité de la Third Offset Strategy (TOS)

La TOS adoptée par le président Obama et renforcée, après lui, par l'administration Trump représente tout à la fois un processus « réflexe » des États-Unis lorsque ceux-ci jugent leur supériorité militaire remise en question et, dans le même temps, une modification sans précédent de la vision géopolitique de Washington à l'égard de ses adversaires et de ses alliés.

La TOS s'inscrit, en effet, dans la longue lignée des stratégies de compensation adoptée par la puissance américaine depuis plus de soixante ans. Ainsi, face à la supériorité conventionnelle de l'Union soviétique sur le théâtre européen, les États-Unis avaient-ils investi massivement dans la modernisation de leur arsenal nucléaire et balistique pour compenser le faible niveau de force et de préparation des pays européens de l'OTAN sur le théâtre centreuropéen. Par la suite, l'Union soviétique rattrapa son retard au point de remettre en question la supériorité militaire des États-Unis

[123] GOVINI, Department of Defense Artificial Intelligence, Big Data, and Cloud Taxonomy, Govini, 2017.

sur le plan nucléaire. C'est alors que les forces armées américaines entamèrent une réflexion profonde sur l'apport des technologies naissantes de l'information, des ordinateurs et réseaux. La seconde stratégie de compensation des États-Unis qui vit le jour, s'appuya sur les avancées dans le domaine des processeurs et des technologies aérospatiales. L'objectif était de réinvestir les forces conventionnelles en faisant de celles-ci le fer de lance de la stratégie américaine contre l'éventualité d'une invasion soviétique de l'Europe. Les nouvelles technologies dans le domaine des munitions guidées de précision, des armes déportées, des contre-mesures électroniques et des senseurs embarqués pour les missions ISR concentrèrent les investissements du Pentagone. L'Initiative de Défense Stratégique (IDS) et le projet « guerre des étoiles » qu'il intégrait en son sein furent les principales vitrines d'une stratégie des moyens beaucoup plus diversifiée destinée à assécher les finances de l'Union soviétique au travers d'une course qualitative aux armements extraordinairement coûteuse.

Toutefois, si elles furent l'expression d'une supériorité technologique américaine écrasante, incontestée et incontestable sur le plan des dispositifs conventionnels dans un rapport de type forces contre forces, tant la victoire de la coalition contre l'Irak lors de l'opération *Desert Storm* que la campagne de bombardement alliée contre le régime serbe de Belgrade à l'occasion de l'opération *Allied Force*, ou encore les campagnes militaires *Enduring Freedom* (Afghanistan, 2001) et *Iraqi Freedom* (Irak, 2003) permirent également à des acteurs résurgents (Russie) et émergents (Chine), ainsi qu'aux adversaires non conventionnels des États-Unis (groupes terroristes islamistes) de comprendre les fragilités d'une organisation de défense principalement axée sur la domination informationnelle et sensible aux pertes humaines. La multiplication de « zones grises » (Ukraine, Syrie, Irak) et le recours à des stratégies de guerre hybride ont considérablement remis en cause la capacité d'intervention des forces armées occidentales et leur aptitude à faire face à certains types de crise. Pour nombre de puissances technologiques, et en premier lieu desquelles figurent les Etats-Unis, la solution aux écueils opérationnels rencontrés par les forces armées dans certaines contingences de crise pourrait passer par

b) Exploration rapide des programmes liés à l'IA

Quelques exemples de programmes suffiront à comprendre l'importance de la mobilisation du DoD en matière d'IA :

- le programme TRACE (*Target Recognition and Adaptation in Contested Environments*) de la DARPA a produit des technologies parmi les plus intéressantes. TRACE fournit, par exemple, un système de reconnaissance automatique des cibles pour aider les pilotes dans le processus de frappe. Plus exactement, le programme TRACE de la DARPA repose sur l'emploi d'algorithmes avancés tels que ceux fournis par la société Deep Learning Analytics LLC à Arlington. L'U.S. Air Force a débloqué un budget de près de 6 millions de dollars américains pour que cette dernière aboutisse au développement d'un système d'appui au ciblage d'objectifs en mouvement ou en reconfiguration. Ce que recherche l'USAF à travers ce programme, c'est la confection d'un système d'intelligence artificielle capable de réduire au minimum les fausses alertes liées à la détection radar de cibles lors d'opération. Le système radar des aéronefs sera donc couplé à un algorithme capable de procéder, sur la base d'entraînements, à l'identification et à la sélection de cibles en temps réel[124] ;

[124] John KELLER, « DARPA TRACE Program Using Advanced Algorithms, Embedded Computing for Radar Target Recognition », *Military Aerospace & Electronics*, 24 juillet 2015, *cf*. https://www.militaryaerospace.com/computers/article/16714226/darpa-trace-program-using-advanced-algorithms-embedded-computing-for-radar-target-recognition

- l'US Air Force développe actuellement un système similaire, appelé Multi-Domain Command and Control. Son but est de regrouper les très nombreuses données provenant d'un large éventail de sources et de créer ainsi une image globale d'une situation. ;
- plus controversé, le Project Maven – précédemment évoqué – est un important projet de renseignement, de surveillance et de reconnaissance conduit sous la houlette du JAIC. Conçu avec l'aide de géants du numérique américains tels que Google, Microsoft et Amazon, il est au cœur d'une polémique depuis qu'une ex-employée de Google qui participa au programme interpella les Nations Unies afin qu'elles bannissent ledit programme ainsi que tout projet susceptible de générer des « robots tueurs »[125]. Utilisant des technologies visuelles assistées par ordinateur, il permet à des analystes de traiter jusqu'à deux ou trois fois plus de données pendant une période donnée[126]. Le système est déjà utilisé dans les opérations de lutte contre Daech ;
- du côté de l'U.S. Army, l'application *Asset Performance Management* de la société Uptake fait l'objet de tests dans la perspective du développement d'une maintenance prévisionnelle sur ses véhicules de combat d'infanterie M-2 Bradley ;
- l'U.S. Army planche encore sur des véhicules de combat de nouvelle génération pouvant facultativement être pilotés ;
- le laboratoire de recherche de l'U.S. Air Force a mis sur pied le programme Skyborg dont le but est de former les pilotes des aéronefs à l'aide d'un système à base d'IA, qui peut en outre être installé dans un aéronef sans pilote[127]. L'objectif du programme Skybord ne réside pas seulement dans l'entraînement des pilotes d'avions de combat. Il vise aussi à permettre au pilote de l'aéronef d'évoluer avec un drone équipé d'un système d'IA. Ce faisant, l'USAF espère réduire le niveau de risque encouru par ses avions pilotés dont la perte, même au niveau unitaire, s'avérerait particulièrement pesante en termes budgétaires. En remplaçant un aéronef piloté – de type F-35 *Lightning* – par un drone plus rudimentaire (à l'instar du XQ-58 *Valkyrie* choisi pour le développement du programme Skyborg), l'USAF espère parvenir à concevoir un multiplicateur de force moins coûteux et plus à même d'opérer dans des environnements risqués[128] ;
- on soulignera encore que la DARPA organise annuellement un défi, le Cyber Grand Challenge. Lors de cette compétition, des machines autonomes s'affrontent entre-elles. Chacune de ces machines est conçue avec des points faibles. Le but pour les participants est de créer des algorithmes d'IA capables de repérer et de combler ces lacunes tout en attaquant leurs adversaires[129] ;
- enfin, l'U.S. Army développe un outil baptisé Macroscope. Ce dernier utilise les données produites par les réseaux sociaux pour mieux comprendre ces environnements.

---

[125] En 2017, après avoir passé quatre ans en tant qu'ingénieure logicielle chez Google en Irlande, Laura Nolan, diplômée du Trinity College Dublin, était affectée au Project Maven du géant californien. Tel qu'on lui avait présenté, ce programme avait pour but d'aider le DoD des États-Unis à améliorer la technologie de reconnaissance vidéo de ses drones. Cependant, rapidement, la jeune femme a commencé à s'inquiéter à propos de l'éthique du projet Maven. Pour cause, il était demandé aux employés affectés à ce programme de développer un système d'intelligence artificielle permettant aux machines de distinguer les humains et les objets instantanément. Laura Nolan a vite compris que le but ultime de ce projet était de permettre à l'armée américaine de cibler et donc de tuer encore plus de personnes dans les zones de guerre comme l'Afghanistan. L'ingénieure logicielle a donc décidé de quitter le Project Maven en 2018, lorsque 3000 employés de Google ont décidé de s'insurger et de signer une pétition contre ce programme meurtrier, forçant ainsi la firme à l'abandonner.

[126] CRS (service de recherche du Congrès américain), « U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress », CRS, 2018.

[127] Valerie INSINNA, « Introducing Skyborg, your New AI Wingman », C4ISRNET, 14 mars 2019.

[128] Harry LYE, « Skyborg: The US Air Force's Future AI Fleet", *Air Force Technology*, 28 août 2019, https://www.airforce-technology.com/features/skyborg-the-us-air-forces-future-ai-fleet

[129] Daniel S. HOADLEY, Nathan J. LUCAS, *Artificial Intelligence and National Security*, CRS, 2018.

3. Les programmes européens

Les États européens et l'UE ont pris très tardivement conscience de l'importance croissante de l'IA et de ses applications, même si des efforts conjugués de recherche ont été entrepris ces dernières années. En vérité, tous les pays membres de l'UE ainsi que la Commission européenne ont adopté des stratégies relatives à cette technologie. De nombreux pays européens, ainsi que l'UE elle-même, ont considérablement augmenté les investissements dans l'IA et mettent en place des structures et des entités pour gérer les opportunités et les défis qu'elle représente. Il n'en demeure pas moins que la recherche dans le domaine de l'IA et des disciplines associées apparaît surtout fracturée en plusieurs initiatives nationales à travers lesquelles les États disposant du capital technologique et industriel souhaitent occuper une place de leader. Par ailleurs, la maîtrise de la *donnée* constitue le maillon faible – voire manquant – de toute entreprise scientifique européenne dans le domaine de l'IA. L'essentiel des échanges de données entre utilisateurs des technologies numériques est contrôlé par des entreprises soit américaines (les fameux GAFAMITIS) soit chinoises. En ce sens, la communauté des utilisateurs européens exporte davantage de *données* qu'elle n'en récolte.

À l'échelle européenne, la mise en place du Fonds européen de la défense et de la coopération structurée permanente pourrait permettre à la R&D sur l'IA appliquée au secteur de la défense de bénéficier d'une dynamique nouvelle. Toutefois, ainsi que nous l'évoquions précédemment, la majorité des initiatives sont envisagées et mises en œuvre au niveau national ou bilatéral. Pire, le Royaume-Uni et la France, qui occupent des places de chefs de file en Europe dans ce domaine, semblent davantage animés par la compétition que par la coopération.

D'une manière générale, l'Europe a accumulé un retard certain par rapport à la Chine et aux États-Unis sur le plan techno-industriel. Dès 2013, un rapport du Sénat français dénonçait cette situation en indiquant que le continent européen était sur le point de devenir « une colonie numérique ». Or, depuis lors, rien ne semble avoir fondamentalement changé. Pour le Dr Laurent Alexandre, du point de vue de l'intelligence artificielle, la France (et l'argument pourrait être étendu à ses partenaires européens) est sur le point de devenir un pays en développement. Elle exporte ses matières premières que sont ses mathématiciens, informaticiens, ingénieurs spécialisés dans l'IA. Dans le même temps, elle importe des biens à haute valeur ajoutée sur le plan technologique issus de pays qui ont su attirer les meilleurs « cerveaux » pour la conception de ces biens. La France, comme l'Europe, exporte encore des données qui sont récoltées, traitées et valorisées par des entreprises qui ne sont ni françaises, ni européennes, mais pour la majeure partie située sur la côte ouest des États-Unis.

a) Le cas du Human Brain Simulation Project

Certes, des initiatives européennes existent mais ont accouché de projets qui furent sujets à des critiques de la part d'experts et de spécialistes. Ainsi en est-il du projet Blue Brain lancé en 2005, qui fut réorienté dans le cadre du Human Brain Simulation Project (HBSP). Choisi par la Commission européenne en 2003 comme l'un des deux programmes phares – FET (Future and Emerging Technology) Flagship Programme – dans le domaine des sciences et des technologies avancées, le HBSP avait initialement pour objectif de parvenir à reproduire le cerveau humain dans le cadre d'un supercalculateur. Mobilisant près de 800 chercheurs répartis à travers 19 pays et incluant 116 sociétés partenaires, le HBSP reçut un budget de 1,2 milliard d'euros sur dix ans. Pourtant, dès le lendemain de son lancement, le programme phare de l'Union européenne allait être au cœur de critiques lorsque pas moins de 800 scientifiques cosignèrent une lettre ouverte à l'adresse de la Commission

européenne dans laquelle ceux-ci remettaient en cause le mode de gouvernance et les orientations d'un projet qu'ils jugeaient alors pharaoniques, pour ne pas dire mégalomane[130].

Il n'est pas ici question de nous attarder sur les multiples crispations et querelles de clocher qui affectent le programme de la Commission européenne, si ce n'est pour retenir les conséquences de ces divergences sur les orientations du HBSP. Au préalable, ceci nous oblige à nous intéresser aux fondements mêmes du projet. L'objectif du HBP est de permettre une meilleure fédération des efforts dans le domaine de la recherche sur le cerveau et les neurosciences. L'IA est supposée constituer l'un des vecteurs ou moyens par lesquels de nouvelles avancées dans la compréhension des mécanismes du cerveau pourraient être envisagées. Le constat opéré par l'Union européenne est que 135 millions d'Européens souffrent de maladies psychiatriques ou neurologiques. De telles pathologies ont des répercussions économiques évidentes puisqu'elles s'accompagneraient d'un coût total situé entre 300 et 700 milliards par an pour l'ensemble des pays européens. De nombreux efforts, souvent mal coordonnés, ont été engagés pour comprendre les origines et les mécanismes de ces pathologies psychiques et neuronales. Les résultats n'ont pas toujours été à la hauteur des investissements. Aujourd'hui, les nouvelles technologies offrent des moyens inédits pour décrypter les masses de données relatives aux gènes et à leurs expressions, la production et la distribution des protéines, les interactions entre protéines, les connexions entre cellules et les liaisons entre les différentes parties du cerveau humain[131]. Le HBSP s'appuie donc sur deux tendances lourdes de la technologie du XXI^{e} siècle. La première est l'accroissement des puissances de calcul des supercalculateurs. Le HBSP, lorsqu'il était encore au stade du projet Blue Brain, a démarré sur la base de supercalculateurs opérant à l'échelle des téraflops. Ce niveau de capacités de calcul permettait de passer de la simulation d'un neurone unique à celle du niveau d'une cellule. Aujourd'hui, les superordinateurs travaillant dans la gamme des pétaflops ont permis de passer de la simulation de la cellule à celle d'un cerveau de rongeur. Des espoirs sont actuellement nourris à l'égard des supercalculateurs à l'échelle des exaflops. Ces derniers n'existent pas encore, alors même qu'ils étaient attendus pour la fin de la décennie qui vient de s'écouler[132].

Le projet HBP souffre de nombreuses incohérences et est aujourd'hui tributaire des divergences entre les principaux leaders du programme. Parmi les détracteurs du programme européen, on citera certains experts perplexes quant aux vertus du HBSP, à commencer par le professeur Jacques Neirynck, professeur honoraire de l'École polytechnique fédérale de Lausanne (EPFL), institut intégré au sein du HBSP. Pour le professeur Neirynck, si le HBSP peut permettre de mieux comprendre le fonctionnement

---

[130] Fabien GOUBET, « Une nouvelle crise secoue le Human Brain Project », *Le Temps*, 21 août 2018, *cf.* https://www.letemps.ch/sciences/une-nouvelle-crise-secoue-human-brain-project

[131] Henry MARKRAMN Karlheinz MEIER, Thomas LIPPERT, Sten GRILLNER, Richard FRACKOWIAK, Stanislas DEHAENE, Alois KNOLL, Haim SOMPOLINSKY, Kris VERSTREKEN, Javier DEFELIPE, Seth GRANT, Jean-Pierre CHANGEUX, Alois SARIA, « Introducing the Human Brain Project », *Procedia Computer Science*, No. 7, 2011, pp. 39-42.

[132] Plusieurs projets de développement de supercalculateurs capables d'opérer à l'échelle de l'exaflops ont vu le jour durant cette dernière décennie. En Europe, ce sont les programmes CRESTA (Collaborative Research into Exascale Systemware, Tools and Applications), DEEP (Dynamical exascale Entry Platform) et Mont-Blanc qui cristallisent la dynamique de recherche autour de cette technologie. Actuellement, le principal effort européen mené sous la houlette de la Commission européenne est le EuroHPC JU (European High-Performance Computing Joint Undertaking). Cette entreprise de recherche formellement établie au sein de l'Union européenne vise le développement d'un supercalculateur à l'échelle de l'exaflop à l'horizon 2022/2023. Les différents membres du programme ont injecté, toutes sources confondues, près d'un milliard d'euros dans ce projet. La contribution financière de la Commission européenne s'élève pour sa part à 486 millions d'euros. Intégré au sein du programme Horizon 2020 de la Commission européenne, l'EuroHPC est divisé en plusieurs jalons de réalisations selon un calendrier prévisionnel serré. Idéalement, le programme doit permettre la conception de deux machines s'approchant de l'échelle des exaflops entre 2019 et 2020. Une capacité de type exaflops est prévue, quant à elle, pour 2022 ou 2023. Enfin, l'EuroHPC envisage de dépasser ce stade en jetant les bases d'un supercalculateur dont les capacités se situeraient au-delà de l'exaflop. Une coordination avec divers centres de recherche dans le domaine du calcul avancé est prévue pour permettre à terme de progresser dans le développement d'un ordinateur hybride conjuguant les capacités à l'échelle de l'exaflop et des technologies du calcul quantique. *Cf.* https://eurohpc-ju.europa.eu

du cerveau, il est en revanche impossible qu'il aboutisse à un système de simulation de celui-ci[133]. D'autres voix se sont élevées pour dénoncer la gabegie d'un programme dont le format s'avère inadapté à la production de résultats. De nombreux observateurs – et parfois même des experts parties prenantes dans le projet – ont fait remarquer que le HBSP présentait davantage de représentants du champ informatique que des neurosciences. Surtout, certains n'ont pas hésité à dénoncer le fait que le HBSP semblait changer de logique. Là où la priorité aurait dû être placée sur les neurosciences et les sciences cognitives (pour une meilleure compréhension des modes de fonctionnement du cerveau) s'est développée une logique essentiellement industrielle. L'accent fut placé sur les technologies de l'information et, plus spécifiquement, sur les six plates-formes destinées à permettre aux chercheurs de décrypter le cerveau. La Commission européenne exige que le HBSP délivre des « produits », selon les termes de Yves Frégnac et Yves Laurent. Dans un article[134] paru dans la revue *Nature*, les deux auteurs, respectivement directeur de recherche au CNRS et directeur du Brain Research à l'institut Max Planck en Allemagne, avaient fustigé les réorientations subtiles, mais néanmoins majeures, imprimées au programme du HBSP. Ils regrettent que l'on ait pensé d'abord aux applications avant même de comprendre le fonctionnement du cerveau.

Si le projet HBSP de l'Union européenne ne résume pas à lui seul la dynamique de recherche existant en Europe dans le domaine des neurosciences et de la simulation sur supercalculateurs des modalités du cerveau humain, il témoigne des difficultés quasi congénitales de l'Union européenne à fédérer des efforts autour d'objectifs « purs » de science fondamentale. Certes, il serait naïf d'imaginer qu'un projet d'une telle ampleur puisse se voir garantir la pérennité d'un financement au seul motif de faire avancer la recherche fondamentale en la matière. Tôt ou tard, la question des applications susceptibles de découler d'une telle recherche se poserait. Toutefois, dans le cas du HBSP, il semble que cette préoccupation de « rentabilité » ait directement prévalu, au risque de mettre en péril les perspectives du projet.

b) Une multitude de programmes nationaux

Le Royaume-Uni a investi un milliard de livres sterling pour devenir un leader mondial de l'IA. Quelques projets et programmes relatifs à l'IA qui sont mis en œuvre dans le secteur de la défense britannique permettent d'entrevoir l'étendue des mesures adoptées :

- le ministère de la Défense et l'administration centrale des communications ont signé un partenariat en matière de défense et de sécurité avec l'*Alan Turing Institute*, un institut spécialisé dans la science des données et l'IA. Ce partenariat est centré sur des projets à long terme, mais il fournit également une plate-forme de formation pour les fonctionnaires ;
- l'État a créé un laboratoire d'IA pour accroître les capacités nationales de défense dans le domaine de l'intelligence artificielle, de l'apprentissage automatique et de la science des données. Ce laboratoire est spécialisé dans les véhicules autonomes, la lutte contre les opérations d'information et l'amélioration des cyberdéfenses ;
- le laboratoire DSTL (*Defence Science and Technology Laboratory*) a organisé plusieurs défis, concours et marathons de programmation (ou *hackathons*) relatifs à l'IA ;
- le DSTL a conçu un système de poursuite radar, *Moonlight*, qui utilise l'apprentissage automatique pour actualiser de façon autonome les informations relatives aux radars ennemis. Ce système fournit en outre des indications et des avertissements aux unités déployées ;

[133] Jacques NEIRYNCK, « L'ordinateur ne peut simuler le cerveau », *Le Temps,* 13 janvier 2019, *cf.* https://www.letemps.ch/opinions/lordinateur-ne-simuler-cerveau

[134] Yves FRÉGNAC, Gilles LAURENT, « Where is the Brain in the Human Brain Project? », *Nature*, Vol. 513, 4 septembre 2014, pp.27-29.

- le projet Nelson de la Royal Navy vise à utiliser l'IA pour concevoir un « cerveau de navire » susceptible d'améliorer les processus décisionnels sur ses navires militaires. Un élément clé est la création d'une plate-forme de données pour l'ensemble de la flotte, qui permet d'avoir accès à toutes les données pertinentes à partir d'interfaces très conviviales ;
- le DSTL a, en collaboration avec des partenaires de l'industrie, développé SAPIENT, un système de capteurs autonome destiné à réduire la charge de travail des opérateurs du renseignement ;
- les systèmes autonomes robotisés ne cessent d'attirer des investissements. En 2018, les forces armées ont testé cinq systèmes de transport sans pilote destinés, par exemple, à acheminer des soldats sur la ligne de front.

La France a investi 1,5 milliard d'euros sur cinq ans dans la R&D consacrée à l'IA et annoncé la création d'instituts interdisciplinaires d'intelligence artificielle qui rassembleront des chercheurs du secteur public et du secteur privé. L'État a par ailleurs reconnu les avantages de l'IA pour le domaine militaire :

- l'agence de l'innovation de défense, qui vient d'être créée, consacrera une part substantielle de son budget de 100 millions d'euros au financement annuel d'activités relatives à l'IA[135] ;
- l'État français a lancé un projet, d'une durée de trois ans, relatif à la collaboration humain-machine pour ses aéronefs de combat et lui a attribué une enveloppe de 30 millions d'euros. L'accent est mis sur les capteurs intelligents/à apprentissage, les systèmes indépendants et les cockpits du futur, ainsi que l'amélioration de la collaboration humain-machine ;
- le projet Artemis vise à mettre au point un système d'IA utilisé pour stocker et gérer le volume énorme de données militaires collectées par la France. Le projet s'appuiera sur les travaux de start-ups, de laboratoires et de petites et moyennes entreprises du secteur civil ;
- le projet Commandement et contrôle des opérations armées a pour but de libérer les commandants opérationnels des tâches répétitives et à faible valeur ajoutée en mettant en place des solutions fondées sur les mégadonnées, l'IA, la réalité virtuelle et d'autres technologies ;
- la start-up parisienne Earthcube a mis au point un logiciel d'analyse des images satellites et a signé quatre contrats avec le ministère français de la Défense.

L'Allemagne a investi 3 milliards d'euros pour la R&D consacrée à l'IA jusqu'en 2025. Le pays s'est également engagé à créer 100 postes universitaires ainsi qu'un réseau de 12 centres de recherche spécialisés dans l'IA. Le développement de l'IA et son adoption par les forces armées semblent encore limités à ce stade. Cela dit, l'Allemagne a proposé à la France un renforcement de leur coopération, notamment dans le domaine de l'IA. L'IA pourrait par exemple jouer un grand rôle dans le projet franco-allemand de système de combat aérien du futur. Ce projet devrait inclure les volets suivants : assistance d'un pilote virtuel ; génération automatique des plans de mission ; adaptation des capteurs à l'environnement ; ajustement de l'interface entre l'humain et la machine en fonction de la charge cognitive du/ de la pilote ; enfin, maintenance prévisionnelle. En août 2018, l'Allemagne a créé une agence semblable à la DARPA qui travaillera sur les technologies de rupture dans le cyberespace. Il est clair que l'IA jouera également un rôle dans ces travaux.

4. La Chine

Comme nous avons eu l'occasion de l'aborder en amont, l'IA est devenue une priorité de premier plan pour les dirigeants chinois. Comme le prévoit le plan de 2017 concernant le développement de l'IA de nouvelle génération, la Chine projette de devenir le leader mondial de cette technologie et de

[135] Wendy R. ANDERSON, Jim TOWNSEND, « As AI Begins to Reshape Defense, Here's How Europe Can Keep Up », Defense One, 18 may 2018.

développer un marché intérieur de l'IA d'une valeur de 150 milliards de dollars d'ici 2030. Les entreprises installées dans le pays jouent déjà un rôle important dans le développement global de l'IA. La Chine investit non seulement sur son sol, mais aussi à l'étranger, ce qui suscite une attention accrue de la part des Alliés OTAN.

Les liens étroits entre le secteur privé et les institutions publiques (comme l'État-parti et les forces armées) facilitent considérablement l'intégration de l'IA dans le secteur de la défense, car les priorités de développement des entreprises sont clairement définies par un processus de coordination descendant. Cette tendance devrait être accentuée par la volonté du président Xi Jinping de mettre l'accent sur la « fusion civilo-militaire »[136].

Selon les analystes, les efforts accomplis par la Chine pour intégrer l'IA dans son paysage militaire sont largement motivés par la perception qu'ont les Chinois de la stratégie des Américains en la matière. Par conséquent, à l'instar de Washington, Pékin s'est focalisé sur le potentiel de l'IA pour améliorer les décisions sur le champ de bataille. Le secteur de la défense chinois travaille en outre beaucoup sur le développement de véhicules autonomes utilisables dans tous les domaines militaires. En 2017, par exemple, une université chinoise ayant des liens avec l'armée a présenté pas moins de 1 000 véhicules aériens Une opérant ensemble sans pilote intégrant l'IA. Il semblerait également que la Chine ait travaillé sur des applications militaires de l'IA dans deux autres domaines : la cybersécurité et les missiles de croisière.

a) réflexion au-delà de la rivalité sino-américaine

La détermination de la Chine à poursuivre ses recherches dans le domaine de l'IA est totale. Le concept d'innovation se situe au cœur des démarches entreprises par la Chine en vue de s'ériger comme la puissance scientifique et technique du XXIe siècle. Xi Jinping, secrétaire général du Parti communiste mais également président de la Commission militaire centrale, a ainsi pu déclarer que, dans une situation caractérisée par une compétition militaire internationale de plus en plus ardue, seuls les innovateurs gagnent[137]. Le président chinois a rappelé à de nombreuses reprises à l'Armée populaire de libération la nécessité pour elle de se transformer en une force de rang mondial pour le milieu de ce siècle. À cette fin, Xi Jinping encourage son institution militaire à tirer le meilleur parti des technologies émergentes et, plus exactement, de l'intelligence artificielle pour respecter le calendrier de modernisation fixé par le pouvoir central. La Chine adopte donc une démarche des plus décomplexée quant à la façon dont elle perçoit sa puissance nationale.

L'intensification de la rivalité entre Washington et Pékin a largement contribué à accélérer le lancement de programmes destinés à accroître la part occupée par les nouvelles technologies au sein du dispositif militaire du pays. Il importe toutefois d'insister sur le fait que ce n'est pas seulement la compétition avec les États-Unis qui explique la détermination de la Chine à investir dans l'IA. La tension existante entre Pékin et Washington ne déterminera pas indéfiniment le cours des relations internationales. Si la Chine investit autant dans les technologies émergentes, les concepts innovants et l'IA, c'est aussi parce que ses élites sont convaincues que l'ontologie même de ces nouvelles technologies affectera durablement les relations stratégiques, les équilibres militaires et la nature de la guerre, au-delà même de cette période de rivalité entre les deux puissances du Pacifique. En d'autres termes, les États-Unis ne représentent pas l'unique paramètre de la politique d'investissement de la

---

[136] Lindsey R. SHEPPARD & al., *Artificial Intelligence and National Security: the Importance of the Ecosystem*, CSIS, novembre 2018.

[137] Déclarations de Xi JINPING telles que rapportées par l'article suivant : Liu GANG [柳刚] et al., Xinhua, « Scientific and Technological Innovation, Towards a Powerful Engine for the World-Class Military » [« 科技创新，迈向世界一流军队的强大引擎 »], Xinhua, 15 septembre 2017, <http://www.gov.cn/xinwen/2017-09/15/content_5225216.htm>, accessed 30 octobre 2019.

Chine dans le domaine des technologies numériques avancées, même si l'idée est clairement assumée de détrôner l'Occident et son monopole technologique[138].

Selon Julien Nocetti, trois jalons permettent de mesurer le niveau des ambitions chinoises en matière technologique. Il s'agit du *volontarisme technologique*, de *l'autosuffisance technologique* et enfin de la *libération entrepreneuriale*.

b) Doper les forces armées par l'IA

L'intelligence artificielle est perçue par les leaders politico-militaires chinois comme une technologie stratégique pour le pays. La crainte du pouvoir de Pékin est de manquer des innovations de rupture susceptibles de renverser l'état actuel de l'équilibre des forces sur le plan international. Il existe, en Chine, cette peur viscérale (ainsi que nous avons pu l'aborder lors du précédent chapitre) de subir un revers ou d'être victime d'une surprise stratégique majeure. Les exhortations multiples du secrétaire général du Parti communiste visant à inciter l'APL à se « doper » à l'IA peuvent surprendre. Elles s'expliquent pourtant au regard de l'histoire de la Chine et du parcours difficile qui fut le sien pour l'acquisition et la maîtrise de technologies clés pour sa souveraineté[139]. Le Plan chinois de développement d'une intelligence artificielle de nouvelle génération (**新一代人工智能**发展规划) appelle au renforcement des modalités d'emploi des technologies d'IA de nouvelle génération pour soutenir les processus de décision du commandement, les inférences militaires lors de simulations ou la conception d'équipements de défense parmi d'autres exemples.

L'APL a choisi d'investir l'essentiel de ses efforts dans le développement de concepts innovants en matière d'opérations « dopées à l'IA ». Des recherches sont notamment conduites dans le domaine des interfaces hybrides humains-machines dans la perspective d'aboutir à des modalités de guerre coopérative de nouvelle génération. En d'autres termes, le projet des forces armées chinoises est d'opérer une transition de la « guerre réseau-centrique » à la « guerre algorithmique » (Algorithm-Centric Warfare), dont les opérations résideraient dans l'emploi massif et ciblé de l'essaimage et le recours aux méthodes de contrôle cognitif. Le concept de « contrôle cognitif » est envisagé par l'APL comme un ensemble de techniques inédites destinées à altérer et manipuler les perceptions de tout adversaire, tant au niveau de ses structures de décision que de son opinion publique[140]. Surtout, l'APL perçoit dans les technologies de l'intelligence artificielle un puissant levier de renversement des équilibres militaires hérités de la domination de l'Occident. L'intelligence artificielle peut incarner cette technologie qui offrira à des acteurs parvenus tardivement sur le terrain des technologies numériques la capacité de procéder à un double saut qualitatif leur permettant de rivaliser avec les puissances technologiques historiques. Le Comité militaire central chinois a ainsi exhorté l'APL à réfléchir à une intégration plus poussée de l'IA en matière de C2. Le projet d'une réduction progressive de la place occupée par l'homme dans les structures de commandement et de

---

[138] Li Bingyan [**李炳彦**], « Major Trends in the New Global Revolution in Military Transformation and Future Warfare Trends » [« **世界新**军事变革大势与未来战争形态 »], Guangming Daily [**光明日**报], 24 janvier 2016. Li Bingyan is a member of the National Security Policy Committee (**国家安全政策委**员会). For an assessment of the PLA's initial reaction to, and assessment of, the Third Offset, see Peter Wood, « Chinese Perceptions of the "Third Offset Strategy" », China Brief, 4 octobre 2016, <https://jamestown.org/program/chinese-perceptions-third-offset-strategy/>, accessed 30 octobre 2019.

[139] Durant de nombreuses décennies, la Chine, en dépit des avancées techniques dont elle sut faire preuve, est longtemps restée dans une dynamique de « rattrapage technologique ». Le secteur spatial chinois en est une illustration. Longtemps dépendante de l'assistance fournie par l'Union soviétique pour son développement, la distanciation intervenue dans les rapports entre Moscou et Pékin avait ralenti de manière notable le développement technologique chinois.

[140] Liu Huiyan [**刘惠燕**], Xiong Wu [**熊武**], Wu Xianliang [**吴**显亮] and Mei Shunliang **梅**顺量], « Several Thoughts on Promoting the Development of Cognitive Domain Operations Equipment Within the Whole Environment » [« **全媒体**环境下推进认知域作战装备发展的几点思考 »], National Defense Science and Technology [**国防科技**], octobre 2018. The authors are affiliated with Unit 61716, the PLA's 311 Base, which is responsible for psychological operations against Taiwan, which has been incorporated into the PLA Strategic Support Force, which also contains its space and cyber capabilities.

contrôle est ainsi clairement évoqué. Des débats existent cependant parmi les instances de réflexion de l'APL sur la place qu'occupera à l'avenir la composante humaine au sein des systèmes de commandement. Comme ailleurs, deux écoles semblent s'opposer : la première milite en faveur d'un retrait à terme de tout intervenant humain dans les processus de conduite des opérations, la seconde se montre plus circonspecte et défend l'idée d'un équilibre entre l'homme et la machine dans la conduite de la guerre. On assiste également à l'émergence d'une réflexion sur les implications éthiques de l'intégration poussée de l'IA au sein des systèmes de force. Ceci n'est pas anodin et mérite d'être souligné. L'apparition d'un tel sujet à l'agenda des réflexions des stratégistes de l'APL confirme une certaine élévation du degré de maturité des responsables chargés des programmes d'intégration de l'IA. On retiendra notamment l'idée, de plus en plus présente et débattue, d'une réglementation internationale des applications militaires de l'IA. Toutefois, de la même façon que les demandes de réforme des grands traités portant sur l'utilisation militaire de l'espace n'ont nullement empêché les démonstrations de force chinoises durant ces quinze dernières années, l'insistance avec laquelle la Chine milite pour l'édification d'un cadre juridique spécifique aux applications militaires de l'IA n'affecte en rien la poursuite des recherches conduites au sein de l'APL. Ainsi est-il par exemple envisagé de poursuivre l'intégration de systèmes d'IA au sein des dispositifs de conscience situationnelle (*Situational Awareness*) et de prise de décision au profit des pilotes des forces aériennes ou encore des commandants de sous-marins. Il a été ainsi rapporté qu'un projet d'intégration d'un système de soutien à la décision à base d'IA est sur le point d'être concrétisé pour les forces nucléaires sous-marines. L'objectif est de réduire la quantité de tâches et surtout la charge psychologique des éléments humains du commandement au sein de ces systèmes d'armes[141].

c) L'IA et l'avenir du commandement

La défaite du champion du jeu de go, Lee Sedol, face au programme AlphaGo de Google en 2016 a constitué un coup de tonnerre au sein de l'élite politico-militaire chinoise. Certains observateurs n'hésitent d'ailleurs pas à qualifier cet événement de véritable « moment Spoutnik » pour Pékin. La capacité pour une machine de renverser au jeu de go une intelligence humaine était une prouesse que même les meilleurs spécialistes de l'IA n'envisageaient pas avant deux décennies. Cet événement confirmait donc une certaine accélération – et non des moindres – de la progression de l'IA sur des terrains jusque-là maîtrisés par l'homme. Sur le plan budgétaire, 2017 confirme la prise de conscience chinoise de l'importance de l'IA : un plan national de 148 milliards de dollars est lancé pour permettre à la Chine de devenir leader dans l'IA à l'horizon 2030. Les stratégistes de l'APL ont cherché à déterminer les applications auxquelles pourraient déboucher à l'avenir les aptitudes des machines basées sur l'apprentissage profond pour le commandement et le contrôle des forces armées. Des concepts innovants ont ainsi fait leur apparition. Lors du salon de l'aéronautique de Zhuhai à l'automne 2018, le China Electronics Technology Group, un conglomérat de défense détenu par l'État, a procédé à la démonstration d'un système de mission à base d'intelligence artificielle pour la conduite des opérations. Le système présenté par le conglomérat avait pour particularité d'être en mesure d'apprendre en parfaite autonomie et d'élaborer un catalogue des modèles de combat. L'avenir d'un tel système dépendra, bien entendu, des données opérationnelles qui l'alimenteront.

La progression de la Chine en matière d'IA à finalité militaire pourrait s'avérer exponentielle dans les décennies à venir. L'APL explore d'ores et déjà une large gamme d'applications de l'IA pour ses systèmes de soutien. On citera, entre autres, l'alerte avancée, le renseignement militaire, les opérations informationnelles (à l'exemple de la cyberdéfense, de la guerre électronique ou de la guerre

[141] Elsa KANIA, « Chinese Sub Commanders May Get AI Help for Decision-Making », *Defense One*, 12 février 2018, <https://www.defenseone.com/ideas/2018/02/chinese-sub-commanders-may-get-ai-helpdecision-making/145906/?oref=d-river>

psychologique), le soutien aux processus de prise de décision et de commandement, ainsi que les systèmes d'armes avancés. La Chine a d'ores et déjà procédé au lancement de satellites d'observation équipés de puces intégrant de l'intelligence artificielle. Cette capacité d'observation renforcée lui permet de disposer d'une imagerie plus performante des zones observées. La Chine entend à l'avenir renforcer ses capacités de surveillance des zones sous-marines à l'aide de dispositifs d'analyse de signaux acoustiques basés sur des réseaux neuronaux artificiels. La possession par la Chine d'une telle capacité, si elle devait se confirmer dans le temps, entraînerait des conséquences non négligeables en matière de dissuasion conventionnelle et nucléaire, dans la mesure où les sous-marins ne seraient plus en mesure d'opérer en totale discrétion.

En marge de ces recherches applicatives, les travaux menés par l'APL portent sur des systèmes avancés de robots (terrestres, navals et sous-marins) et de drones avec IA embarquée. Une attention particulière est attachée au développement des robots navals et submersibles spécialisés dans la surveillance en milieu maritime. Les forces aériennes de l'APL, quant à elles, mènent des travaux dans le domaine de l'acquisition de systèmes de drones de combat et d'essaims droniques. On mentionnera encore les activités de recherche et développement de l'industrie de défense chinoise dans le champ des missiles, et plus particulièrement des missiles de croisière. On connaît depuis de nombreuses années les progrès réalisés par la Chine dans le domaine des missiles hypersoniques et hypermanœuvrants. Les avancées réalisées par l'APL sont de nature à impacter les équilibres militaires régionaux et globaux dans la mesure où de telles capacités sont susceptibles de remettre en question la protection – certes, relatives – des systèmes de défense antimissile déployés tant par les États-Unis qu'au niveau de l'Alliance atlantique. Toutefois, au-delà de cette modernisation, la Chine semble avoir amorcé des avancées notables dans la fabrication de systèmes de croisière disposant de plus d'autonomie avec de l'intelligence artificielle embarquée. L'objectif est de permettre au système d'armes de disposer d'une plus grande autonomie opérationnelle et décisionnelle qui puisse décharger les états-majors tout en permettant à ceux-ci de disposer d'une capacité de contrôle en temps réel du système. Les futurs missiles de croisière chinois pourraient donc embarquer à leur bord une IA dotant le système d'armes d'un certain niveau de cognition lui permettant d'opérer en mode « tire-et-oublie » tout en s'adaptant à l'évolution de l'environnement de bataille.

d) Le développement d'un écosystème d'innovation

Les autorités politico-militaires chinoises semblent faire preuve d'une certaine lucidité quant aux perspectives de modernisation de leur appareil militaire en vue de faire entrer celui-ci dans le XXI^{e} siècle. Si la recherche de défense chinoise multiplie les concepts innovants en matière d'armements, il lui reste à développer un écosystème de recherche adapté à la réalisation de ruptures technologiques. Il va sans dire que la base industrielle et technologique de défense chinoise dépend de près ou de loin, directement ou indirectement, des orientations gouvernementales du Parti communiste. Qu'il s'agisse des centres de recherche, des laboratoires ou des principaux groupes industriels que compte le pays, des liens particulièrement étroits existaient entre ces diverses entités et le pouvoir central de Pékin. En 2014, toutefois, le Premier ministre Li Keqiang prononçait à Tianjin un discours qui marquait une rupture claire avec le mutisme du gouvernement chinois en matière de politique numérique. Désormais, Pékin se montrait favorable aux investissements privés de fonds capital-risque et au développement de start-ups et d'incubateurs dans le domaine de l'IA[142]. C'était là,

[142] Kai-Fu LEE, *AI Superpowers: China, Silicon Valley, and the New World Order*, Boston/New York, Houghton Mifflin Harcourt, 2018.

ni plus ni moins, une véritable garantie offerte par l'État pour l'évolution des rapports entre le secteur public et le secteur privé en matière de recherche technologique[143].

Confirmation de cet élan nouveau, le plan « Made in China 2015 » présentait désormais la substance des ambitions technologique de la Chine. Il s'agissait plus précisément de soutenir l'élévation en gamme de l'économie chinoise en abaissant la dépendance du pays aux technologies issues de l'étranger et en investissant dans des secteurs stratégiques comme l'intelligence artificielle et la robotique. L'approche de Pékin est de tirer parti du marché intérieur pour dynamiser l'innovation en matière de science et de technologie et, surtout, de garantir l'autonomie stratégique de la Chine dans le secteur des technologies de rupture.

Le développement d'un écosystème de recherche consacré à l'IA en Chine suppose aussi la capacité d'attirer les compétences venues de l'étranger. En ce sens, la Chine est en train d'amorcer une guerre des talents en multipliant les politiques attractives destinées à faire venir sur son sol les meilleurs experts et scientifiques spécialisés dans le numérique et l'IA[144]. Non seulement la Chine aspire littéralement les données provenant des utilisateurs de ses technologies mais elle concentre de manière croissante les meilleures expertises que compte la planète dans le domaine de l'intelligence artificielle.

La Chine présente un double atout sur ses rivaux occidentaux dans sa quête pour la maîtrise de l'IA[145]. Le premier est d'avoir amorcé une stratégie de fusion civilo-militaire en s'appuyant sur la participation des géants technologiques que compte le pays à l'instar de Baidu, Alibaba et Tencent. Ces grands groupes ont accepté de participer à une large gamme de projets de coopération avec des instances de recherche militaires à d'évidentes fins militaires. Un second avantage – et non des moindres – dont dispose la Chine sur ses compétiteurs est l'absence totale de retenue des autorités politiques du pays à collecter les données de ses quelque 800 millions d'internautes. Les masses de données récoltées permettent d'alimenter des bases de données colossales dont les contenus vont permettre l'éducation des systèmes d'IA. Si, en retour, les consommateurs chinois profitent de la finesse des systèmes d'IA pour leurs applications quotidiennes, l'accroissement des performances des intelligences artificielles grâce aux données récoltées auprès des utilisateurs permet à Pékin de renforcer ses moyens de surveillance sur l'ensemble de la population. La Chine se transforme, pour reprendre les termes de Julien Nocetti, en une « Arabie saoudite de la donnée » tant elle dispose d'une réserve colossale d'informations qui, selon les experts, constituera la véritable réserve stratégique du XXI^{e} siècle. Cette information d'importance nous amène à une dimension souvent méconnue de la politique technologique chinoise : l'influence que Pékin entend exercer sur l'instauration future de normes et de standards dans le secteur de l'IA. Il est hors de question, en effet, pour la Chine de se voir imposer à l'échelle internationale des normes en matière d'IA auxquelles elle n'aurait pas contribué. Aussi, la diplomatie chinoise se montre-t-elle particulièrement active et contribue-t-elle de manière intense aux travaux conduits dans le cadre de l'Organisation internationale de normalisation et du *Partnership on AI*, un consortium composé d'industriels présidé par les principaux groupes du numérique étatsunien. En 2018, le consortium dont il est question a accepté d'intégrer Baidu parmi ses membres.

Comment l'ensemble de ces dynamiques et politiques se mettent-elles en place ? Quels sont les résultats concrets des efforts entrepris dans le cadre d'un rapprochement entre le monde civil et le monde militaire en Chine ? On peut, à cet égard, prendre en exemple l'Université de Tsinghua qui se

[143] Julien NOCETTI, *Intelligence artificielle et politique internationale : les impacts d'une rupture technologique*, Paris, Institut français des relations internationales, Études de l'IFRI, novembre 2019, p. 18.
[144] Julien NOCETTI, « La Chine, superpuissance numérique ? », dans Thierry DE MONTBRIAL, Dominique DAVID, *RAMSES 2019. Un monde sans boussole ?*, Paris, Institut français des relations internationales /Dunod, 2018, pp. 124 – 129.
[145] *Ibid.*, Paris, Institut français des relations internationales, Études de l'IFRI, novembre 2019, p. 18.

destine à devenir une sorte de « MIT à la chinoise ». La collaboration instituée entre la recherche civile et militaire au sein de cette institution est essentiellement destinée à des projets militaires dans le domaine de l'intelligence artificielle. Disposant de fonds provenant de la Commission des sciences et technologies du CMC, l'Université de Tsinghua s'est engagée sur de nombreux programmes militaires parmi lesquels figurent des projets de collaboration synergique humain-machine. L'un des projets à long terme de cette université est de développer un laboratoire entièrement consacré à la supervision des programmes à finalité militaire. Autre exemple de la politique scientifique insufflée par le pouvoir en faveur de l'IA militaire, le Centre d'innovation et de fusion civilo-militaire en matière d'IA (人工智能军民融合创新中心) basé à Tianjin. La spécificité de ce centre est d'avoir été édifié à proximité immédiate du Centre national de supercalcul qui abrite le supercalculateur Tianhe-3, premier prototype de supercalculateur capable d'opérer à l'échelle exaflopique (soit un milliard de milliards d'opérations à la seconde, virgule flottante).

5. La Russie

Le président Vladimir Poutine a déclaré : « L'IA est l'avenir […]. Quiconque occupera la première place dans ce domaine deviendra le maître du monde ». Bien que la Russie accuse encore un certain retard par rapport aux États-Unis et à la Chine, elle a montré sa volonté de rattraper ses concurrents, tout au moins dans certains domaines. Cela dit, alors que les entreprises chinoises et américaines consacrent des milliards de dollars à l'IA, le secteur privé russe n'y investit que 700 millions de roubles environ (moins de 11 millions de dollars)[146].

Le ministère de la Défense, ainsi que certains acteurs de l'industrie de la défense, jouent un rôle prédominant au regard de l'IA. Tout d'abord, la Commission de l'industrie militaire russe souhaite que 30 % de son équipement militaire soient contrôlables à distance à l'horizon 2025[147]. Dans le cadre de cet effort, le gouvernement russe a créé une fondation pour la recherche de pointe – le pendant russe de la DARPA –, dont le budget annuel se situe aux alentours de 4 milliards de roubles (environ 62 millions de dollars). Cette fondation s'est concentrée jusqu'ici sur les technologies imitant la pensée humaine, l'analyse des données et l'assimilation de nouvelles connaissances. Elle a également défini quatre grands axes de développement de l'IA : la reconnaissance des images, la reconnaissance de la parole, le contrôle des systèmes militaires autonomes, et enfin le soutien pendant le cycle de vie des systèmes d'armes. Les industries russes intègrent l'IA dans les systèmes d'armes, en particulier les systèmes autonomes robotisés. Le groupe Kalachnikov aurait mis au point un véhicule terrestre commandé par intelligence artificielle, capable d'identifier et d'engager des cibles de façon autonome (IISS, 2018). La société KRET, spécialisée dans les technologies radio-électroniques, travaillerait sur « des systèmes sans pilote capables de prendre en toute indépendance des décisions en essaim »[148]. Par ailleurs, l'armée de l'air russe a annoncé la conception de missiles guidés à l'aide de l'IA. Il se pourrait également, selon les analystes, que les technologies civiles de reconnaissance des images et de la parole – dont le développement est bien avancé en Russie – soient intégrées dans les opérations d'information russes. Il convient toutefois de noter que les projets ambitieux de la Russie dans le domaine de l'IA pourraient être mis à mal par les problèmes structurels du pays, par exemple la faiblesse de l'industrie technologique et la baisse des budgets de la Défense.

[146] Michael C. HOROWITZ & al., *Strategic Competition in an Era of Artificial Intelligence*, Center for a New American Security, 2018.
[147] Greg ALLEN, Taniel CHAN, *Artificial Intelligence and National Security*, Belfer Center for Science and International Affairs, 2017.
[148] IISS, « Big data, artificial intelligence and defence », *The Military Balance 2018*, IISS, 2018.

Il apparaît que, dans son ensemble, le budget consacré à la recherche militaire dans le domaine de l'intelligence artificielle reste bien en deçà des niveaux atteints par d'autres puissances militaires. Il existe, en vérité, un différentiel de plusieurs générations entre les montants colossaux investis à divers degrés par les forces armées des États-Unis et les budgets consacrés par l'armée russe. Il n'en demeure pas moins que l'intelligence artificielle intéresse au plus haut point la Russie pour ses besoins opérationnels. Habituées à intervenir dans des zones urbanisées, à haute densité de population, sur des terrains où les armées occidentales redoutent d'intervenir, les forces armées de la Fédération de Russie tireraient grand parti des potentialités militaires de l'IA. Depuis le milieu des années 1990, la Russie est intervenue sur des théâtres particulièrement dimensionnants à l'instar de la Tchétchénie, du Daghestan, de la Géorgie, de l'Ukraine ou de la Syrie. Depuis 2008, un processus de modernisation de ses forces et de ses équipements a été engagé afin d'élever le niveau qualitatif et de préparation de ses troupes. Le spectre d'une rivalité croissante avec les États-Unis est toujours présent et la crainte de voir la Chine la déborder sur le plan des équipements est un scénario qui gagne en probabilité avec le temps.

La Russie marque un intérêt certain pour le développement de solutions fondées sur l'intelligence artificielle avancée à destination de l'ensemble de ses besoins opérationnels. Ceux-ci incluent la robotique militaire (pour le déploiement en zones 3D), l'autonomie, l'apprentissage machine pour la collecte des données et leur traitement en informations exploitables par le commandement, l'aide à la navigation et à la conduite sur des terrains difficiles d'accès, etc. Depuis 2008, il est vrai, le pays est parvenu à procéder à un certain nombre de bonds qualitatifs remarqués pour certaines capacités clés tels que des systèmes aériens inhabités, des dispositifs C4ISR et des systèmes balistiques ou de croisière.

C'est cependant à partir de 2017 et les déclarations du président Poutine que la dynamique russe dans le secteur de l'IA semble avoir été engagée de manière concrète. En 2018, le ministère du Développement économique tenait une table-ronde consacrée à la problématique dans le cadre du forum militaro-technique ARMY-2018. Toutefois, depuis cet événement, aucune autre rencontre n'a formellement été organisée sur ce sujet. L'une des particularités des recherches russes menées dans le domaine de l'IA tient au fait que celles-ci sont essentiellement orientées vers des besoins en robotique militaire. Le compendium encyclopédique militaire russe présente ainsi le concept de « robot de combat » : un dispositif technique multifonctionnel caractérisé par un comportement anthropomorphique, capables d'effectuer partiellement ou intégralement des fonctions habituellement réalisées par un humain lors de missions de combat.

Le ministère russe de la Défense classe les projets de robots de combat en trois générations :

- une première génération regroupe les robots intégrant des logiciels et des programmes de contrôle à distance pouvant opérer uniquement dans un environnement organisé (ce qui exclut en principe le champ de bataille) ;
- une seconde génération de robots feront montre d'un comportement plus adaptatif et disposeront à échéance d'organes sensoriels artificiels les rendant capables d'évoluer dans des environnements plus aléatoires ;
- enfin, une troisième génération de robots sera incarnée par des systèmes dotés d'intelligence artificielle embarquée leur assurant une autonomie de déploiement.

Avant même que ne surviennent les percées dans l'IA et l'apprentissage machine de ces dix dernières années, les forces armées russes avaient accordé une attention croissante à la problématique de l'intégration de la robotique au sein de ses systèmes de forces. En 2000, le ministère russe de la Défense avait ainsi adopté un programme cible d'intégration intitulé « Robotisation des armes et de l'équipement militaire – 2015 ». Ledit programme avait pour objectif de favoriser une dynamique de R&D destinée à produire et tester de manière expérimentale des prototypes de robots terrestres.

Cependant, en dépit de l'effort qui semblait amorcé au niveau politique, aucun démonstrateur technologique financé par ce programme ne vit le jour, ce qui conduit à un arrêt brutal de la R&D en la matière.

Il fallut attendre l'année 2015 pour voir ressurgir une initiative nouvelle. Au mois de septembre 2015, le ministère russe de la Défense adopta un programme intitulé « Création d'une robotique militaire avancée d'ici 2025 avec prévisions à l'horizon 2030 ». Ce nouveau programme a pour but de développer des véhicules inhabités de type « robots » destinés à intervenir dans des environnements de combat « complexes ». L'ambition affichée par la Russie est d'intégrer 30 % de systèmes télépilotés au sein de son arsenal d'ici 2025. C'est dans la foulée de ces mesures que le président Poutine put signer le 16 décembre 2015 un décret instituant un centre national pour le développement des technologiques et des éléments fondamentaux de la robotique. Ce centre est appelé à s'insérer au sein d'un Fonds pour la recherche avancée (FRA).

En décembre 2016, le gouvernement russe adopta une Stratégie de développement scientifique et technologique pour la Fédération. Parmi les priorités affichées figurent la transition vers un environnement numérique plus poussé, les technologies de conception intelligentes, les systèmes robotiques, les méthodes et matériels innovants de fabrication, le développement de systèmes pour l'analyse des bases de données, l'apprentissage machine et l'intelligence artificielle. La Fédération de Russie procéda également à une redéfinition des cadres de sa recherche militaire en accouchant d'une structure nouvelle se présentant comme suit :

- une Commission du ministère de la Défense consacrée au développement des systèmes robotiques à finalité militaire. Cette entité est présidée par le ministre de la Défense en personne. Ladite commission est chargée du développement de procédures et d'organisations pour la conception des systèmes robotiques. Son travail consiste principalement à favoriser la cohérence des recherches au travers d'une unification et d'une coordination entre les différents départements en charge des recherches et applications dans ce domaine ;
- un Département principal pour la recherche et le soutien aux technologies avancées est également institué (institution référencée sous l'acronyme GUNID). Le GUNID fait partie intégrante du ministère de la Défense. Il intervient au titre de principal spécificateur des systèmes robotiques militaires et participe au développement de procédures et de protocoles communs ;
- un Centre principal de recherche et d'expérimentation des systèmes robotiques du ministère de la Défense (le plus souvent référencé sous la désignation anglaise de Main Research and Testing Robotics Centre – MRTRC) a également été créé. Il est l'un des organismes les plus secrets au sein des forces armées russes. Il est d'ailleurs très rare que le MRTRC rende compte de ses réalisations. Parmi les quelques projets aboutis connus qui sont issus du MRTRC figure, semble-t-il, le réseau d'information et de mesure de la Marine pour l'observation de l'Arctique. Le mot d'ordre du directeur du MRTRC, Sergueï Popov, est « Smart, Small, Many and Inexpensive » ;
- s'ajoute à cet ensemble institutionnel un Fonds pour la recherche avancée (déjà évoqué plus haut). L'une des principales difficultés pour la Russie dans la recherche technologique de pointe fut de faire en sorte que les budgets promis soient réellement convertis en dépenses effectives par les centres de recherche et les laboratoires. C'est afin de palier à ce problème que le président Poutine semble avoir particulièrement insisté sur la nécessité de créer un tel Fonds dont le rôle se rapproche fortement de celui joué par la DARPA aux États-Unis. Ce fonds de recherche constituera sans nul doute l'agence de prédilection pour la conduite politique de nombreux programmes tant au sein de l'Armée de terre que de la Marine ;

- enfin, un Technopole de l'innovation militaire (généralement référencé sous l'acronyme ERA) a vu le jour. Cet organisme regroupe une douzaine d'entreprises scientifiques et technologiques du ministère de la Défense, soit près de 600 membres de personnel travaillant sur des programmes à destination de la Défense. Ce technopole d'innovation mène des projets de recherche et de développement prioritaires dans des domaines tels que les systèmes de télécommunication et d'information, les systèmes automatisés et robotiques, l'intelligence artificielle, la conception numérisée, les systèmes de sécurité ou encore les nanotechnologies et les nanomatériaux.

Les entreprises investies dans la R&D en matière de défense en Russie tentent, au travers de ces initiatives gouvernementales, de suivre le rythme de l'innovation de leurs homologues américaines, israéliennes et sud-coréennes, notamment dans la conception de systèmes robotiques avec intelligence artificielle embarquée. De nombreux spécialistes, cependant, estiment que la Russie devrait au minimum doubler ses efforts afin de figurer dans le peloton de tête de cette nouvelle course qualitative aux armements. Le choix délibéré de la Russie d'aborder l'IA au travers des systèmes robotiques et des technologies de l'autonomie présente cependant un certain nombre de risques dont celui de la difficile intégration au sein de systèmes qui pourraient présenter des limites en termes de déployabilité et de mise en œuvre. La priorité qu'accordent les multiples instances investies dans la recherche sur l'IA et les dispositifs robotiques ainsi que leur crainte d'être dépassées dans le cadre d'un environnement hypercompétitif ont conduit à un effacement quasi complet des débats portant sur les enjeux éthiques de tels systèmes dans un cadre militaire. Les seules questions abordées à propos des programmes en cours sont leur compatibilité avec les perspectives d'évolution de la Convention sur certaines armes conventionnelles dans le cadre de l'ONU.

6. Au niveau de l'OTAN

L'IA n'a jamais figuré à l'ordre du jour d'un sommet de l'OTAN. Néanmoins, les responsables politiques et militaires de l'Alliance ont assisté à des exposés sur cette question lorsque le Conseil de l'Atlantique Nord et le Comité militaire ont consacré leur journée informelle de 2018 aux technologies de rupture, dont l'IA. Cela dit, plusieurs entités au sein de l'OTAN ont lancé des activités relatives à l'IA ou inclus cette technologie dans leurs autres activités au cours des dernières années :

- l'Organisation pour la science et la technologie (STO) a consacré deux de ses trois thèmes scientifiques et technologiques à l'utilisation de l'IA et des mégadonnées dans la prise de décision et l'autonomie. Cette démarche ainsi que la tenue d'une réunion rassemblant de nombreux spécialistes sur le premier thème ont entraîné une augmentation des activités relatives à l'IA dans le programme de travail collaboratif de la STO, par exemple des travaux sur le contrôle humain significatif, la cyberdéfense faisant appel à l'IA et l'utilisation de l'IA dans le domaine de l'information ;
- le Commandement allié Transformation (ACT) organise un certain nombre d'événements consacrés aux opportunités et aux défis de l'IA, par exemple lors des forums OTAN-industrie ainsi que des conférences internationales sur le développement et l'expérimentation de concepts ;
- l'Agence d'information et de communication de l'OTAN (NCIA) a fait de l'IA le sujet central de son colloque de 2018 sur l'assurance de l'information. En novembre 2018, la NCIA a également organisé un marathon de programmation, baptisé « Hackathon for Good », dont le but était de mettre au point des outils d'analyse de mégadonnées, de visualisation de données et d'apprentissage automatique pour faire face à des opérations d'information. L'IA sera également l'un des nombreux sujets à l'ordre du jour de la conférence de l'industrie ;

- le Groupe consultatif industriel de l'OTAN (NIAG) s'est lui aussi investi dans le domaine de l'IA. Il a récemment produit deux études sur le sujet : l'une sur l'utilisation par l'OTAN des mégadonnées, l'autre sur l'impact de l'autonomie sur les plans et les opérations de l'OTAN (Blunt, Riley et Richter, 2018) ;
- le programme OTAN pour la science au service de la paix et de la sécurité (SPS) a explicitement sollicité des propositions concernant l'usage de l'IA dans la lutte contre le terrorisme dans son appel à propositions de 2017.

Sur le plan pratique, l'Alliance a déjà fait appel aux mégadonnées et à l'apprentissage automatique, par exemple pour éliminer les doublons ou les copies redondantes de jeux de données recueillis par sa mission en Afghanistan ou figurant dans les fichiers journaux de détection de comportements anormaux. Des produits et des services fondés sur l'IA ont également été utilisés lors d'un exercice de réponse en cas de catastrophe organisé en 2018 avec la Serbie, partenaire de l'OTAN.

Conclusion partielle Les quelques déclarations péremptoires de chefs d'État et de gouvernement semblent attester de l'existence d'une course à l'intelligence artificielle entre les nations les plus avancées sur les plans scientifique et technologique. Plusieurs observations peuvent être faites à ce stade.

Premièrement, force est de constater le manque de coordination des efforts entrepris par les différents États. La course à l'IA semble s'inscrire dans une logique essentiellement nationale, avec un niveau de coopération très faible sur le plan multilatéral. Les quelques projets engagés en commun s'avèrent de portée limitée, quand ils ne sont pas les otages de rivalités nationales ou institutionnelles.

Deuxièmement, on ne peut nier la prépondérance de la rivalité entre les États-Unis et la Chine dans le développement des systèmes d'intelligence artificielle. Il est évident que cette compétition à l'échelle globale semble s'inscrire dans une lutte géopolitique. C'est aux États-Unis et en Chine que l'on assiste à la mise en place de véritables stratégies, accompagnées d'institutions spécifiques et de budgets d'une ampleur avec laquelle les États européens ne parviennent pas pour l'heure à rivaliser.

Troisièmement, la finalité des efforts engagés tant par les États-Unis que la Chine ou la Russie est clairement militaire. Washington, Pékin et Moscou abordent l'IA sous l'angle quasi exclusif des applications militaires susceptibles de découler des avancées technologiques qui lui sont associées. La conviction selon laquelle l'IA s'avérera un facteur de différenciation déterminant pour l'avenir des équilibres militaires est fermement ancrée dans les représentations qui animent les politiques étrangères et scientifiques de ces acteurs. L'Europe semble, quant à elle, se cantonner pour l'essentiel à un rôle de régulateur.

VI. Enjeux éthiques et juridiques

L'une des caractéristiques fondamentales de l'innovation technologique depuis l'avènement de l'ère industrielle est d'avoir sans cesse produit des effets qui ont dépassé l'entendement humain sur les conséquences à long terme. On pourrait penser, au regard d'une certaine actualité, que la question de l'expansion de l'intelligence artificielle échappe à ce drame de la créativité technique de l'homme. Des voix – et non des moindres – se sont élevées pour dénoncer les dérives susceptibles de découler d'une prolifération incontrôlée de l'IA. Des appels réitérés en faveur de l'adoption d'un moratoire sur les développements à venir de l'IA ont fait florès ces dernières années comme nous l'avons vu. De telles mises en garde émanant de scientifiques et technologues réputés ne sont pas rares. Elles font d'ailleurs figure de contre-balancier aux promesses faites par les techno-évangélistes sur les bienfaits futurs – et encore mal définis ou incertains – des innovations technologiques actuelles et à venir.

Pourtant, les avertissements ou appels à la cessation des recherches relatives à l'IA et à ses technologies associées ne sont pas sans présenter un certain nombre de difficultés et témoignent de nombreux biais, quand ils ne s'appuient pas sur des incompréhensions profondes des technologies concernées et de ses enjeux.

L'un des premiers travers constatés parmi les discours prônant une interdiction de l'IA dans le champ militaire est celui qui consiste à regrouper en un seul et même ensemble des systèmes d'armes de nature et de conception différentes. Il n'est pas rare, en effet, d'observer une confusion regrettable (et parfois, entretenue à dessein) entre intelligence artificielle, armes autonomes (incluant souvent les armements semi-autonomes, téléguidés) et drones. L'objectif des détracteurs de toutes ces technologies de défense vise précisément à favoriser l'émergence d'un amalgame entre, d'une part, ce qui existe (et est dénoncé) et, d'autre part, ce qui relève au mieux d'une orientation technoscientifique, sinon du fantasme pur (dans l'état actuel du développement technologique)[149]. La simplification à outrance du débat sur l'intégration des dernières avancées technologiques du numérique au sein des dispositifs militaires des nations industrialisées favorise trop souvent les prises de position extrêmes marquées par l'agitation de symboles[150] et la confusion des genres : la plus regrettable étant l'indifférenciation entre considérations éthiques et juridiques.

En guise de prologue pour ce chapitre, il nous semble utile de clarifier un certain nombre de considérations. Une première considération consiste à distinguer les technologies de l'IA des « systèmes d'armes létaux autonomes » et des drones armés. En l'état actuel du développement technologique, aucun de ces systèmes ne pose de réels nouveaux enjeux en matière politique, militaire, éthique ou juridique, et ceci pour une raison simple : leur appellation s'avère tronquée et ne correspond pas à la réalité technique profonde de ces systèmes. Ainsi les systèmes d'IA spécifiques auxquels recourent les forces armées ne sont-ils pas totalement indépendants de la décision humaine. Leur conception est humaine et résulte de l'ingénierie humaine. Il n'existe pas, à ce jour, de systèmes d'IA capables de créer d'autres systèmes d'IA, et la perspective d'un IA forte (ou IA générale) n'est encore qu'un scénario d'évolution parmi d'autres (cette hypothèse a été l'objet d'un précédent chapitre). L'existence d'une IA forte supposerait que l'on ait affaire à un système informatique parfaitement autonome capable de s'affranchir de la décision humaine. Ce n'est aujourd'hui pas le cas. De semblables mises en garde de vocabulaire s'imposent en ce qui concerne les systèmes d'armes létaux autonomes. L'autonomie évoquée à propos de tels dispositifs désigne surtout des

[149] Jean-Baptiste JEANGÈNE-VILMER, « Légalité et légitimité des drones armés », *Politique étrangère*, 2013/3, pp. 119-132.

[150] Ainsi, Joseph HENROTIN regrette-il que le drone soit devenu « la figure aérienne du mal », alors même que les modalités qui président à son recours s'inscrivent dans des pratiques de la guerre qui ont toujours été adoptées et consenties par les États et leurs peuples. *Cf.* Joseph HENROTIN, « Le drone : figure aérienne du mal ? », *Défense & Sécurité Internationale Hors-Série*, numéro 30, juin – juillet 2013.

« automatismes » sous supervision humaine. La notion d'autonomie appliquée aux robots militaires porte à confusion. Ce terme est aujourd'hui employé pour décrire un panel de caractéristiques allant du simple automatisme à un hypothétique système d'IA qui serait doté d'une conscience. Même dans le cas d'assistants tels que Google Assistant, Alexa ou Siri, c'est la capacité d'accès de ces systèmes à une somme gigantesque de connaissances et de données qui génère l'illusion de l'intelligence ou de la conscience[151]. Comme le rappelle fort pertinemment Dominique Lambert, « une exigence éthique fondamentale est le respect de la vérité. Sans elle, quelque chose du rapport au réel est perdu. Or, ici, un problème de vérité peut se poser. En effet, lorsque l'on évoque les robots autonomes (qu'ils soient physiques ou électroniques, d'ailleurs), on fait comme si la simulation ou la représentation de capacités et de performances humaines par des machines autorisait une sorte d'identification de l'humain et du robot. Mais cela ne va pas de soi, même si le vocabulaire employé peut laisser à penser le contraire, puisqu'on dit que le système " pense ", " décide ", " juge ", etc.[152] » Dans les faits, l'intelligence artificielle, au travers de ses algorithmes, constitue une réduction du réel, de la même manière qu'en mathématiques on distingue les structures de leurs représentations ou qu'en physique on établit une distinction entre le modèle du réel et le réel lui-même. Un algorithme est donc une réduction, une approximation du mode de fonctionnement de l'intelligence humaine. Peut-être même qu'un algorithme n'imite en rien les structures cognitives humaines, puisque la compréhension de ces dernières par la science comporte encore bien des zones d'ombre. Et Dominique Lambert de s'interroger : pourquoi même devrions-nous, par conséquence, imaginer que les robots (ou l'IA) devraient être identiques ou semblables à l'humain ? Cette question a été abordée, nous nous en souvenons, dans le cadre de notre réflexion sur le transmachinisme.

Par ailleurs, dans le domaine militaire, la décision de tir dans le cas du déploiement de tels systèmes appartient à un opérateur humain et ne saurait être confiée au système d'armes en lui-même. Autrement dit, les systèmes d'armes « parfaitement » autonomes n'existent pas pour l'heure et ne posent donc pas d'enjeux nouveaux sur les plans moral ou juridique. Quant aux drones armés, notamment employés pour des assassinats ciblés dans le cadre de la lutte anti-terroriste menée par certains pays (dont les États-Unis), le principal problème provient d'une confusion entre la *fin politique* et *l'instrument* au service de cette fin politique. Clarifions donc : si l'usage des drones armés ne pose aucune difficulté d'ordre juridique particulière (pas plus que n'importe quel autre système conventionnel[153], comme les missiles de croisière), la politique dans laquelle cette technologie s'insère peut, pour sa part, offrir matière à discussions et à controverses. Précisons. Dans le cadre d'un conflit armé, l'utilisation de drones armés n'est légale que si elle se conforme aux règles du droit international humanitaire. En d'autres termes, le recours au drone armé ne s'entend que dans le cadre d'un conflit armé au sens du DIH. En dehors du cadre d'un conflit armé tel que défini par le DIH, le recours au drone armé est illégal et soumis au droit commun.

D'une façon plus générale, il semble essentiel de souligner que ce n'est pas toujours la nature de l'arme qui rend celle-ci illégale. Certes, certains types d'armement sont interdits par le droit international. Ainsi en va-t-il par exemple des mines anti-personnel. D'une manière générale, du point de vue du droit, lorsque ce n'est pas la nature de l'arme qui fait l'objet d'une interdiction, c'est son emploi qui est soumis à une régularisation ou à un certain nombre de limitations. Poser la question de la légalité

[151] Joël MORILLON, « L'autonomie des robots terrestres est-elle pour demain ? », *Les Cahiers de la Revue Défense Nationale*, Dossier Autonomie et létalité en robotique militaire, Institut des hautes études de défense nationale (IHEDN) et le Centre de recherche des écoles de Saint-Cyr Coëtquidan (CREC Saint-Cyr), 2018, p. 57.

[152] Dominique LAMBERT, « Éthique et autonomie : la place irréductible de l'humain », *Revue Défense Nationale*, Dossier « L'intelligence artificielle et ses enjeux pour la Défense », mai 2019.

[153] Le champ nucléaire, du fait de la nature du niveau de destruction pouvant être atteint, se situe dans un référentiel juridique et éthique propre (mais pas sans lien avec le conventionnel).

de certains dispositifs tels que les systèmes d'IA ou les SALA ne consiste pas à rechercher dans ce qui constituerait « l'essence » de l'armement considéré un élément conduisant à son illégalité. La légalité ou, surtout, l'illégalité d'une arme résulte d'une convention, d'une décision, d'un contrat passé entre partenaires en vue de la qualification d'un armement auquel on renonce à recourir. Cet aspect ne doit jamais être perdu de vue. La question sur le plan de l'éthique est tout autre.

Tant l'intelligence artificielle transposée dans le contexte militaire que les systèmes d'armes létaux autonomes sont avant tout l'objet d'un procès d'intention de la part de ceux qui les dénoncent. Ceci ne signifie pas que leur rôle au sein des organisations militaires ne pose pas de questions sur les modalités de conduite de la guerre post-moderne. Mais elles doivent être correctement formulées et porter sur les spécificités techniques réelles de ce qui les constituent.

1. Le cas des SALA

Ce sont plus de 70 pays qui se sont réunis à l'Office des Nations unies à Genève entre le 27 et le 31 août 2018 pour débattre du statut des systèmes d'armes létaux autonomes dans le cadre de la Convention sur certaines armes conventionnelles (CCAC). C'était alors la sixième fois depuis 2014 que de telles réunions avaient rassemblé les représentants diplomatiques et les experts de plusieurs pays. Pourtant les avancées – pour peu que l'on s'accorde à reconnaître leur existence – furent rares et se limitèrent à l'identification des enjeux techniques, politiques, moraux, juridiques ou encore à la circonscription des différents concepts – souvent équivoques – véhiculés lors de ces discussions.

L'une des principales difficultés rencontrées par les représentants des États, les négociateurs et les délégués provenant des organisations non -gouvernementales (ONG) concernées résida précisément dans l'élaboration d'un cadre « normatif », destiné à s'appliquer à une technologie qui, pour l'heure, est purement et simplement inexistante. Le constat, que nul ne viendrait nier, selon lequel de nombreux pays sont à l'œuvre pour le développement de systèmes d'armes létaux autonomes (SALA) ne constitue en rien une base solide pour l'élaboration de règles qui viendraient à s'appliquer à ceux-ci le moment venu, et ce pour au moins deux raisons.

La première raison est que l'élaboration d'une norme – surtout dans le domaine épineux qu'est celui du contrôle des armements – exige la maîtrise d'un ensemble des données techniques extrêmement précises sur les principes de conception et de mise en œuvre des armements. Or, la transmission d'information par les États à propos des percées techniques issues de leurs recherches est parcellaire.

La seconde raison est qu'il est difficile de déterminer le moment à partir duquel de tels systèmes seront considérés comme existants et déployés. Envisager la définition a priori d'un cadre réglementaire, adapté à des technologies dont on ignore pour l'heure l'horizon d'émergence tant elles dépendent de facteurs fluctuants liés à l'avancement de la recherche et à la compétence des ingénieurs, relève de la pure gesticulation incantatoire. Le concept d'autonomie figure par ailleurs, on le sait, au cœur même des débats portant précisément sur la « réalité » de telles armes.

Or il n'existe pas de définition univoque de ce que serait la caractéristique essentielle d'un SALA : son autonomie. L'autonomie est souvent considérée selon une approche graduelle et non comme un seuil à partir duquel un système d'armes serait considéré comme récipiendaire d'une capacité décisionnelle indépendante. Les débats qui se sont tenus à Genève à la fin du mois d'août ont d'ailleurs révélé les doutes exprimés par certains représentants sur l'utilité d'un exercice normatif destiné à réguler une réalité technique pour le moins fuyante.

En l'état actuel du développement technique, la perspective d'une arme susceptible de s'affranchir de toute forme de supervision humaine ne présente aucune utilité militaire.

En effet, l'autonomie parfaite d'un SALA impliquerait que nous ayons affaire à une arme non seulement en mesure de s'assigner ou de modifier des objectifs sans validation humaine préalable (perspective d'ores et déjà préoccupante) mais, qui plus est, capable de s'affranchir du cadre de règles qu'un homme ou qu'une instance humaine lui aurait confié. En d'autres termes, si le respect d'un dispositif normatif applicable aux SALA ne dépend pas uniquement de l'homme pour son respect et sa mise en œuvre, quelle serait la plus-value réelle d'un semblable dispositif sur le plan de la pratique stratégique ? À moins d'appeler à la table des négociations un représentant… de la communauté des SALA.

a) Effets pervers

La mobilisation énergique des acteurs investis dans les négociations visant l'adaptation de la CCAC aux systèmes d'armes létaux autonomes pourrait, par ailleurs, générer des effets opposés à l'objectif recherché qui réside, rappelons-le, dans la limitation d'une éventuelle course aux armements entre nations disposant de la base industrielle et technologique requise pour la confection de ce type d'armes.

Une analogie intéressante pourrait être faite entre les négociations actuelles sur le CCAC et les positions des différents protagonistes principaux de ce que fut en son temps la Conférence navale de Londres de 1930 (et dont l'objectif était une réduction globale des forces navales des grandes puissances européennes). Plutôt que de parvenir à une diminution des niveaux de forces en présence (notamment en Méditerranée) et à une réduction des risques d'éclatement d'un affrontement (notamment entre la France, d'une part, et l'Allemagne et l'Italie, d'autre part), les pourparlers de la Conférence révélèrent l'intransigeance des positions de chacun des protagonistes et mirent en avant la nécessité pour les différents acteurs de parvenir à atteindre des niveaux de forces dont il serait difficile d'obtenir a posteriori la réduction.

Pas plus que ladite Conférence de Londres n'évita l'éclatement du second conflit mondial, les discussions autour des SALA n'empêcheront pas la recherche en matière de technologie de poursuivre sur sa lancée (et pas seulement dans le domaine de la défense) afin de doter certains systèmes d'une autonomie croissante en matière décisionnelle et de redéfinition de missions. Ces négociations contribueront même – certes, involontairement – à inciter les acteurs industriels à accélérer les projets dans ce domaine, avant qu'un hypothétique dispositif normatif ne vienne contraindre les activités du secteur. En d'autres termes, l'enjeu pour la science appliquée et l'industrie sera de normaliser la technologie avant que celle-ci ne soit normée.

Cette perspective se révèle particulièrement inquiétante quand on sait l'opposition qui existe au cours des discussions de la CCAC entre les partisans d'un moratoire sur ce type d'armement du futur, d'une part, et les opposants à toute forme d'interdiction à leur endroit, d'autre part. Les États-Unis, Israël, l'Australie, la Russie et la Corée du Sud ont clairement manifesté leurs réticences à toute limitation ou tout bannissement des SALA.

Les États-Unis figurent parmi les pays les plus déterminés à faire avancer la science fondamentale et la technologie en matière d'IA combinée à des systèmes d'armes. Les principaux axes de recherche visent précisément à accroître les capacités de perception, de traitement et de raisonnement des machines. Les SALA, et plus spécifiquement la militarisation de l'IA, forment d'ores et déjà le cœur de la course aux armements que se livrent les principales puissances militaires planétaires. Xi Jinping, le président chinois, n'avait-il pas lui-même déclaré qu'il espérait voir son pays devenir la première puissance militaire mondiale grâce à la maîtrise de l'IA?

b) Les risques d'un possible moratoire

On le voit, la dénonciation *ab initio* de l'illégalité ou de l'illicéité de certaines armes, avant même que celles-ci n'existent, posent des difficultés importantes au regard de la validation du raisonnement juridique[154]. Certes, on pourrait arguer qu'il serait raisonnable d'appliquer un moratoire à certaines catégories d'armements avant même que ceux-ci n'aient engendré d'effets dévastateurs. En effet, il s'agirait sans doute d'une transposition dans le domaine de la défense du principe dit de « précaution ». Oui, mais… Une interdiction aussi générale que théorique sur le statut d'un armement qui n'existe pas encore et dont on ne connaît pas les modalités de fonctionnement peut s'avérer contre-productive et aboutir à des situations que l'on souhaitait initialement éviter. Toute forme d'interdiction générale, abstraite ou théorique peut très rapidement s'avérer inapplicable dans la pratique. En outre, face à une telle conjoncture, la réaction de certains États désireux d'avancer dans la mise en œuvre de tels armements le jour venu de leur avènement consisterait précisément à multiplier les arguments justifiant la non-applicabilité du moratoire au seul prétexte que les critères contenus dans ce dernier ne correspondent plus à la réalité du système dénoncé[155].

Les défenseurs d'un moratoire « de principe » des SALA (sans réellement savoir ce qu'il convient de placer, aujourd'hui, au sein de cette catégorie assez floue) estiment qu'en vertu de la « spécificité » de l'arme (spécificité restant tout entière à démontrer pour l'heure), un dispositif réglementaire particulier mériterait d'être développé en vue de leur interdiction.

Les partisans d'une prétendue incompatibilité « essentielle » entre les SALA et le droit international humanitaire s'appuient sur une vision tout à la fois tronquée et schizophrénique du DIH. Le droit international humanitaire est, en effet, un cadre juridique évolutif et pragmatique. Il peut s'appliquer aux nouvelles technologies puisque l'article 36 du premier protocole additionnel des Conventions de Genève de 1949 impose aux Etats d'évaluer la licéité de nouvelles armes[156]. En effet, l'article 36 dispose que :

> « *Dans l'étude, la mise au point, l'acquisition ou l'adoption d'une nouvelle arme, de nouveaux moyens ou d'une nouvelle méthode de guerre, une Haute Partie contractante a l'obligation de déterminer si l'emploi en serait interdit, dans certaines circonstances ou en toutes circonstances, par les dispositions du présent Protocole ou par toute autre règle du droit international applicable à cette Haute Partie contractante* ».

Il semble désormais établi que, dès lors qu'ils seront mis en œuvre (ce qui n'est pas encore le cas du fait de leur inexistence), les SALA n'échapperont pas aux dispositions du DIH. Toute forme de moratoire conduirait à une situation dans laquelle la rivalité entre les États dotés de ce type de technologies favoriserait le principe de dissimulation et de soustraction à des mesures légales : des pratiques déjà mises en œuvre par les Etats en temps normal.

---

[154] Nathalie DURHIN, « Systèmes d'armes létaux autonomes : ne pas mélanger juridique et éthique », *Les Cahiers de la Revue Défense Nationale*, 2019, p. 167.

[155] Il convient de rappeler que ce fut là l'une des difficultés rencontrées par le traité sur les forces conventionnelles en Europe (FCE), dont un grand nombre de dispositions s'étaient avérées inadaptées tant à la transformation du contexte stratégique issu de la fin de la guerre froide et de la modification des alliances subséquentes qu'à l'évolution des armements et de leurs niveaux de déploiement.

[156] Ainsi, notons que le terme « arme » de l'article 36 renvoie à tout dispositif offensif ou défensif spécialement conçu pour blesser, tuer, endommager ou neutraliser des personnes et/ou des biens.

2. IA et application du droit humanitaire

En l'état actuel du développement des technologies, les systèmes de drones (même armés) déployés sur des théâtres d'opération ne semblent pas poser d'interrogations nouvelles sur le plan de leur conformité au droit international et, en particulier, sur le plan du droit international humanitaire, si et seulement si leur recours intervient dans le cadre d'un conflit armé tel que défini par le DIH. Seul le cadre de leur recours (comme lors de campagnes d'élimination ciblée menées par les États-Unis) suscite des questions réelles sur le plan légal.

Il nous reste, cependant, à envisager l'un des aspects les plus discutés de l'apport futur de l'IA dans les opérations militaires : la capacité d'un système d'intelligence artificielle militaire à faciliter l'application du droit international humanitaire. Nombre de commentateurs font état, de manière quelque peu expéditive, de l'incapacité d'un système d'IA à garantir le respect des dispositions du DIH dans le cadre d'un conflit armé, en général, et d'une opération militaire, en particulier. Dans la conduite de leurs opérations, les forces armées sont continuellement amenées à distinguer entre combattants et civils. Cette nécessaire distinction, imposée par le DIH, résulte d'une vérification obtenue par une procédure complexe d'analyse et de contrôle du théâtre d'opération. Il n'est pas rare que les forces de l'adversaire se dissimulent expressément parmi la population civile afin d'empêcher l'intervention des forces armées. En outre, dans le cadre d'un conflit armé non international, seule la participation effective aux hostilités permet de distinguer le combattant de la population civile (cette dernière étant couverte par la protection offerte par le DIH). Il est souvent argué que les robots, même aidés par l'IA, ne pourraient procéder à la distinction effective entre civils et combattants. C'est là une approximation regrettable de la réalité à laquelle sont confrontés au quotidien les forces armées déployés sur des théâtres d'opération. Il n'est pas une action des forces armées expéditionnaires qui ne fasse pas l'objet d'un passage au crible des senseurs pour une analyse la plus précise possible de la répartition entre combattants et populations civiles. Le commandement en charge de missions est continuellement confronté à des dilemmes entourant la question du respect de critères – souvent cumulatifs – sur le statut des individus présents sur une zone d'opération : y a-t-il port d'armes ? Y a-t-il déplacement à l'aide d'un véhicule militaire ? etc. Autant de questions qui aident les forces armées à s'assurer des forces adverses en présence mais dont la résolution peut s'avérer complexe au point de conduire à l'annulation d'opérations.

Un robot doté d'IA peut-il aider à de telles vérifications ? Dans une certaine mesure, les opérations militaires contemporaines conduites par les nations industrialisées recourent d'ores et déjà aux technologies des capteurs embarqués par des systèmes robotiques. La véritable question qui est ici désormais posée est de savoir si un système robotique létal, doté d'IA, pourrait être en mesure de garantir en conditions opérationnelles le respect des principales dispositions du DIH.

Selon plusieurs experts, il serait possible de codifier en données algorithmiques les principaux critères autorisant l'emploi de la force armée selon le DIH. Avant d'explorer plus en détails cette affirmation, il est utile d'évoquer plusieurs qualités que pourraient présenter les SALA. La première est la facilité avec laquelle de tels systèmes pourraient se rapprocher au plus près de la cible en vue d'analyser les conditions de l'environnement et la nature des protagonistes en présence. Pour autant qu'une doctrine d'emploi en prévoie le cas de figure, un SALA pourrait être chargé de l'analyse de la cible au plus près de celle-ci. Un second atout du SALA résiderait dans l'absence de toute forme de stress dans le chef du système robotique déployé.

Qu'en serait-il du respect des principes de distinction et de proportionnalité par un SALA ? Un SALA doté d'IA pourrait garantir selon plusieurs experts le respect du principe de distinction civil/combattant sur un théâtre d'opération. Les données recueillies par les capteurs embarqués, conjuguées aux éléments d'observation de senseurs téléportés (drones), contribueraient à une analyse plus fine et

plus fiable de la répartition entre combattants et civils sur une zone d'intervention. Cela ne signifie pas que le SALA serait en mesure de prendre la décision du tir en autonomie. Un SALA disposerait, certes, d'une autonomie dans les fonctions critiques de sélection et d'attaques de cibles mais intégrerait une chaîne de commandement et de contrôle (C2) qui le soumettrait à la décision humaine pour l'enclenchement du tir[157]. Semblables affirmations exigent toutefois la plus grande prudence et, sans nul doute, la plus grande retenue. L'interprétation des données issues de senseurs – aussi diversifiés et complexes soient-ils – ne suffit pas à apprécier une situation à l'aune du principe du respect des critères du DIH. Surtout, les contingences de crise actuelles présentent des conditions de terrain extrêmement complexes. Dans la plupart des pays en conflit, une grande partie de la population peut être armée. Cela n'en fait pas pour autant des combattants effectifs. Par ailleurs, comment différencier dans de telles zones de guerre des véhicules militaires de véhicules civils ou de transport ? La plupart des véhicules employés par les combattants sur le terrain sont en tous points similaires à la majorité des véhicules circulant dans ces zones.

Lors d'interventions critiques, des systèmes de type SALA équipés d'IA se révéleraient, toujours aux dires des promoteurs des SALA, plus fiables qu'un combattant humain. L'insensibilité des robots aux sentiments et passions ressentis par les humains qualifieraient même davantage les SALA à des missions critiques tout en assurant le respect des règles du DIH. Du reste, les capteurs les plus divers embarqués par de tels systèmes, eux-mêmes reliés à un réseau informatique pour l'analyse en temps (quasi) réel des données de la situation critique, les conduiraient à pouvoir agir (toujours sous supervision humaine) avec une promptitude et une précision que ne pourraient jamais égaler les combattants humains. Les scénarios les plus divers d'une intervention de SALA sur le champ de bataille ou sur des théâtres d'opération peuvent être envisagés de manière théorique. Sur le plan pratique, des vérifications et certifications devront être mises en œuvre afin de valider l'opérationnalité d'un SALA et le respect par celui-ci du DIH. Mais comment procéder à de telles certifications ? Il convient avant tout de préciser que ce seront aux responsables humaines – politiques et militaires – d'amorcer les réflexions nécessaires à l'encadrement doctrinal de tels systèmes lorsque ceux-ci seront produits et jugés opérationnels. Des procédures rigoureuses devront être conçues qui préciseront, entre autres, les cas dans lesquels la décision pourra être déléguée à l'échelon le plus bas (c'est-à-dire le robot) et les contextes dans lesquels la décision d'engagement du tir devra être réservée à des échelons d'un niveau supérieur, de nature humaine. Des SALA équipés d'IA ne constituent pas, en soi, une menace au principe du contrôle des opérations par des humains. L'absence de contrôle n'est pas une fatalité. Certains commentateurs ont d'ailleurs suggéré de faire passer à des SALA le test d'Arkin avant tout déploiement sur un théâtre d'opération. Le test d'Arkin est l'équivalent du test de Turing pour les systèmes robotiques dotés d'une autonomie relative. Son résultat doit permettre de décider si, à conditions égales, un robot peut démontrer qu'il peut respecter le DIH tout aussi bien (ou mieux) que ne le ferait un être humain[158]. Pour Ronald Arkin, auteur du test, les robots démontreront à l'avenir des capacités bien supérieures à celles de l'homme pour la conduite d'opérations militaires. Il n'en demeure pas que le degré d'avancement technologique des SALA n'impliquera en rien un abandon par l'homme des décisions à la faveur de quelque « libre arbitre » de la machine. Les organisations politico-militaires entendront disposer de moyens soumis à leur contrôle. La technologie, autrement dit, ne pourra jamais dédouaner les différents acteurs intervenant dans la chaîne logistique et de commandement qui préside à la conduite d'une opération faisant intervenir des SALA.

[157] Nathalie DURHIN, « Systèmes d'armes létaux autonomes : ne pas mélanger juridique et éthique », *Les Cahiers de la Revue de Défense Nationale*, 2018, p. 172.

[158] Ronald C. ARKIN, « The Case for Ethical Autonomy in Unmanned Systems », *cf.* https://www.cc.gatech.edu/ai/robot-lab/online-publications/Arkin_ethical_autonomous_systems_final.pdf

Il va de soi que l'ensemble des considérations et hypothèses qui viennent d'être évoquées n'offrent aucune garantie sur leur effectivité et encore moins de dimension pratique. Concrètement, il semble très difficile d'imaginer que dans un avenir plus ou moins proche, non seulement la technologie permettra à la machine d'opérer une évaluation des situations à propos de leur conformité avec le DIH, mais encore que les nations s'accorderont pour le déploiement de tels systèmes. Surtout, la caractéristique première de toute opération militaire est son irréductible imprévisibilité. « Aucun plan ne survit au contact de l'ennemi » affirmait Erwin Rommel. Cet enseignement se révèle encore plus pertinent dans le contexte des crises actuelles marquées par la diversification des acteurs, des moyens, des objectifs de guerre et des visées politiques. La planification des opérations, même à l'aide d'IA ou en ayant recours à des futurs hypothétiques SALA pour leur mise en œuvre, ne peut s'affranchir d'une évaluation permanente de la situation compte tenu des facultés d'adaptation de l'ennemi.

3. Une réflexion éthique qui doit être distinguée du droit

Le débat sur le statut des SALA au sein des organisations militaires a trop souvent tendance à confondre deux ordres de questionnements : celui de la légalité par rapport au droit international (et en particulier le DIH) et celui de de la moralité, c'est-à-dire de la compatibilité d'une technologie avec un système de valeurs.

Sur un plan strictement juridique, rien ne s'oppose au développement d'une arme nouvelle, même autonome, dès lors que l'on peut s'assurer qu'elle ne contrevient pas au respect du droit international humanitaire. La conception d'une arme nouvelle peut intégrer une limitation des risques liés à l'imprédictibilité et aux défaillances techniques. Il semble opportun de rappeler que le droit définit ce qui est autorisé et ce qui est défendu et non ce qui relève du bien ou du mal. Le bien et le mal sont des principes posés par la morale. Le raisonnement juridique s'appuie sur une norme posée par l'autorité publique et interroge le comportement des personnes physiques et morales au regard de la règle de droit dans un lieu donné et en un temps donné. L'éthique relève d'un autre registre. Elle définit ce qui est acceptable sur le plan social, culturel, politique, voire moral. L'éthique peut grandement changer selon les époques et les lieux. Le droit peut, dans certains cas, correspondre à une règle éthique sans toutefois se confondre avec elle. On distingue, en général, plusieurs courants dans l'éthique. On parle « d'éthique de conviction » lorsqu'elle érige des règles, par nature relatives, en principes absolus. On parle « d'éthique de responsabilité » lorsqu'il s'agit d'appréhender l'acceptabilité d'une décision, d'un développement scientifique, d'une évolution technologique, etc. Le droit suppose donc la seule sujétion d'un comportement ou d'un fait à une norme générale et abstraite ; la réflexion éthique vise la recherche du « bien » ou d'un « mieux » au niveau de la société.

Comme cela a été dit, la réflexion éthique peut appuyer la norme, mais elle ne peut jamais se confondre avec elle. Or, on peut regretter l'entretien d'une confusion entre l'argumentation juridique autour des SALA et la réflexion éthique. Cette confusion semble de prime abord perceptible au travers des arguments mêmes du CICR lorsque celui-ci évoque la responsabilité de l'État lorsqu'il envisage la conception ou le déploiement de tels systèmes :

> « *Le droit international humanitaire exige de ceux qui planifient, décident et mènent des attaques qu'ils effectuent certains jugements pour respecter les règles lors du lancement d'une attaque. Des considérations éthiques vont de pair avec cette exigence : elles requièrent le maintien d'une intervention et d'une intention humaines dans les décisions de recours à la force* ».[159]

[159] CICR, « Un nouveau pas vers l'imposition de limites à l'autonomie des systèmes d'armes », Déclaration du CICR du 12 avril 2018.

Une telle confusion entre le droit et l'éthique n'est à première vue pas heureuse mais elle n'est qu'apparente. Il y a, d'un côté, le droit applicable et de l'autre l'éthique qui détermine quand et comment l'humain recourt à la force. Il existe donc toujours un dialogue entre le domaine du droit et celui de l'éthique dans le champ militaire. L'éthique doit avant tout être envisagée comme un éclairage pour le droit.

Ce dialogue entre le droit et l'éthique nous conduit à aborder l'argument de la *déshumanisation*. L'idée sous-jacente à ce raisonnement consiste à affirmer que le recours à des SALA déshumaniserait l'usage létal de la force. Donner la possibilité à une machine d'ouvrir le feu constituerait une incompatibilité avec la « dignité humaine ». Il existe incontestablement un problème d'ordre éthique à une telle conjecture. Selon certains défenseurs des SALA, cette hypothèse ne s'opposerait toutefois nullement à quelque norme du droit international, encore moins du droit international humanitaire. Rien, en effet, *stricto sensu*, dans le DIH ne proscrirait l'usage de la force létale par une machine. Un aspect de cette réalité juridique est toutefois souvent occulté : il s'agit de l'argument selon lequel l'éthique peut – et doit – servir ici de base à la formulation et à l'interprétation du DIH. Souvent qualifiés de nouveaux « abolitionnistes » (autrement dit les partisans d'une interdiction pure et simple des systèmes d'armes létaux autonomes et autres systèmes automatisés militaires), les détracteurs des SALA insistent sur la nécessité de permettre au droit, et plus spécifiquement, aux règles du droit international humanitaire de tendre vers un principe éthique. Beaucoup ont cru voir dans ce dialogue entre le droit et l'éthique un mélange des genres venu s'ajouter une indifférenciation entre SALA et drones armés. Le qualificatif souvent usité pour ces deux familles de systèmes – appelés *robots tueurs* – est d'ailleurs regrettable. Mais on se félicitera de constater que les critiques formulées à propos des généralisations opérées par les tenants d'un moratoire sur les « robots tueurs » aient poussé les organisateurs de la campagne « Stop Killer Robots » à clarifier leur position en prenant soin de préciser que leur démarche visant à la promulgation de moratoires sur certains types de technologies militaires ne devait nullement constituer un frein à la recherche technologique ni même à la robotique en tant que telle.

À ce stade de notre étude de la question du statut des SALA et du principe d'autonomie parmi les « machines », il nous faut insister sur le fait qu'une réflexion se doit d'être poursuivie sur la place des systèmes télépilotés ou automatisés dans le champ militaire. Cependant, que constatons-nous ? Qu'une pareille réflexion ne semble pas avoir été amorcée en son temps notamment à propos du statut des missiles de croisière. Or les caractéristiques de l'arme (à l'exception de son caractère consommable) ne semblaient pas lui permettre d'échapper au débat qui fait actuellement rage à propos des systèmes « autonomes ». C'est là un point sur lequel des éclaircissements devraient être apportés par les « abolitionnistes ».

4. Conclusion partielle

Comme nous pouvons le percevoir, le lien souvent établi – avec quelque facilité – entre le recours aux drones armés et la place de l'intelligence artificielle sur des théâtres d'opérations extérieures constituent deux thématiques qui, sans toutefois être hermétiquement séparées sur le plan de la réflexion théorique, ont fait l'objet d'une confusion regrettable sur le plan pratique. L'emploi des drones armés dans le cadre de campagnes d'éliminations ciblées pose, il est vrai, des questions de premier ordre qui relèvent essentiellement du droit international et de la politique étrangère. La généralisation du recours aux drones armés dans la lutte extérieure contre le terrorisme pourrait constituer la première étape d'une transformation inquiétante du statut de l'espace aérien en droit international. Il est utile de rappeler que la souveraineté nationale sur les espaces aériens n'a pas toujours constitué un principe général immuable. En 1910, lors de la première conférence internationale pour la navigation aérienne, la France était partisane d'une liberté totale de la

circulation aérienne tandis que la Grande-Bretagne faisait valoir la nécessité de maintenir une souveraineté nationale sur les espaces aériens surplombant les territoires terrestres. Une décennie plus tard, Londres opéra un revirement à 180 degrés et milita en faveur d'une liberté généralisée de la circulation aérienne. La diplomatie britannique se heurta alors aux Etats qui, soucieux de préserver leurs compagnies aériennes nationales et inquiètes d'éventuelles incursions inamicales dans leurs cieux, batailllèrent pour le maintien du statut des espaces aériens nationaux. La conduite d'opérations par drones armés, même dans le cadre de l'élimination de cibles individuelles, constitue une véritable remise en cause du statut des espaces aériens nationaux.

La tolérance à l'endroit de l'emploi du drone armé peut-il conduire de manière imperceptible à l'acceptabilité des SALA lorsque ceux-ci apparaîtront et se généraliseront ? C'est là, à dire vrai, la question sous-jacente figurant en filigrane du débat sur les conséquences du drone. La réponse à cette question, comme nous l'avons vu, est loin de faire l'unanimité. Pour plusieurs experts, la tentation pourrait être grande de faire évoluer les drones actuels vers des systèmes plus complexes, toujours plus autonomes, au point de pouvoir un jour s'affranchir de l'intervention humaine pour la conduite d'opérations létales. Cette hypothèse nous amène à une nouvelle question : la disponibilité et la maturité des technologies en matière d'autonomie impliquera-t-elle automatiquement l'abandon par l'homme de son pouvoir de décision dans le recours à la force et l'exercice de la violence armée légitime ? Si la réponse logique à cette interrogation est « non », il serait faire de cécité intellectuelle de ne pas percevoir dans le continuum du développement technique portant le drone au statut de SALA le risque d'un glissement politique permettant à terme la délégation de la décision de frappes à une IA, soit que celle-ci opère des machines à distance, soit qu'elle soit elle-même à bord de cette machine. A dire vrai, la solution aux multiples dilemmes que semblent poser les hypothèses futures de recours au SALA serait d'une simplicité déconcertante : il pourrait être purement et simplement affirmé qu'une décision humaine précise et circonstanciée précède le plus immédiatement la frappe. Bien sûr, il s'agit là d'une solution de nature juridique. Or, la place des SALA au sein de nos organisations militaires et de nos stratégies relève du domaine politique, voire de la sphère sociologique. Pour paraphraser Robert Oppenheimer lorsqu'il s'exprimait à propos des raisons qui ont conduit à l'émergence des travaux sur la bombe thermonucléaire, sans doute que le « délice technique » d'implique la solution des SALA risquera-t-elle de lever les inhibitions qui jusque-là auront empêché leur déploiement.

Conclusion générale : l'IA peut-elle transformer la guerre ?

Nous voici arrivés au terme de notre parcours sur les perspectives d'application militaire de l'intelligence artificielle. Le cheminement que nous avons suivi est nécessairement incomplet et ne saurait prétendre à la moindre exhaustivité. Nous avons pu mesurer les incertitudes qui entourent, aujourd'hui comme hier, le sens même de cette expression désormais usuelle qu'est devenue l'intelligence artificielle. Du statut de champ d'étude, l'IA a progressivement désigné des dispositifs, des machines supposées reproduire les fonctions cognitives humaines et permettre une meilleure connaissance du cerveau humain et de l'intelligence biologique. Mais l'IA est aussi et surtout devenue le centre de gravité de visions diverses (politiques, socioéconomiques, militaires), de scénarios d'anticipation, de fantasmes. Elle a acquis le statut de levier stratégique : elle peut ainsi propulser un État à un statut de puissance sans que ce dernier n'ait à passer par les différents paliers d'une telle progression. Cependant, un constat s'impose aujourd'hui : on lit tout et n'importe quoi sur l'IA[160]. Cette observation est d'autant plus évidente lorsque nous abordons les débouchés militaires potentiels de l'IA. Dans une certaine mesure, nous pourrions dire que le discours sur l'IA, dès lors qu'il est véhiculé par les milieux politiques, en est venu à s'affranchir de toute considération matérielle et technique pour ne retenir que les promesses que ses symboles comportent ou, inversement, les menaces que son expansion pourrait receler.

Affirmer que les principales percées, c'est-à-dire celles récemment intervenues dans le domaine de l'apprentissage machine et des réseaux de neurones profonds, découlent des investissements et de la recherche conduite par les grands groupes privés du numérique est un raccourci mensonger, tant il néglige l'impulsion originelle qui fut celle des milieux scientifiques militaires. L'IA, qu'on la considère comme un champ d'étude ou une technologie, doit son existence à l'institution militaire. De la même façon, il serait inexact de dire que ce sont des fonds privés qui ont permis d'aboutir aux résultats aujourd'hui obtenus. Les aboutissements de la recherche appliquée dans le domaine de l'IA ont bénéficié de nombreux investissements publics dans la recherche fondamentale.

Technologie supposée offrir des réponses aux problèmes rencontrés par les forces armées dans le cadre des opérations présentes et à venir, l'intelligence artificielle et ses multiples déclinaisons sont également porteuses d'interrogations quasi infinies sur le rapport de l'homme, du politique et de la société à la guerre. L'IA place nos institutions de défense face à des défis nouveaux qui, loin d'être insurmontables, exigent qu'un débat soit clairement posé non seulement à propos des méthodes de combat de demain mais plus encore sur l'adaptation de nos mécanismes décisionnels en cas de crise. Ce point se révèle plus aigu encore dans le contexte de la dissuasion nucléaire et de la tentation de confier à des systèmes d'IA certaines clés d'activation dans des scénarios d'échanges nucléaires spécifiques (nous visons ici plus exactement la garantie d'une capacité de seconde frappe).

L'état actuel de la technologie dans le domaine de l'IA n'est pas de nature à vider de toute substance les socles de la pensée stratégique. Le recours par les armées aux outils et aux machines est une caractéristique des plus communes de l'histoire militaire. La recherche du meilleur armement, des meilleures technologies (et nous pouvons inclure dans cet ensemble les processus décisionnels à la base des actions militaires) ou encore de l'avancée matérielle décisive a tout autant structuré les opérations militaires que la discipline, l'esprit de corps, le dévouement et le sacrifice. Les technologies de l'autonomie (bien que ce terme soit imparfait), l'intelligence artificielle et toutes ses ramifications ascendantes (technologies dont elle est issue) ou descendantes (technologies auxquelles elle a pu donner naissance) n'affectent nullement la nature même de la guerre, même si elle modifie sa pratique

[160] Miguel BENASAYAG, *La tyrannie des algorithmes – Conversation avec Régis Meyran*, Paris, Textuel, coll. Conversations pour demain, 2019.

par les nations qui disposent d'une certaine allonge technologique. Tant que l'IA ne produira pas elle-même ses propres systèmes d'IA, ses propres algorithmes ou ses propres armements, rien de ce qui pourra résulter de l'emploi de l'IA dans nos organes de décision ou nos dispositifs opérationnels ne sera en mesure de transformer la guerre.

La science-fiction, littéraire ou cinématographique, a pu, il est vrai, véhiculer un certain nombre d'approches apocalyptiques liées à l'émergence de l'IA. Nous ne connaissons que trop les référents habituellement convoqués par ceux qui interrogent les experts à propos des risques que pourraient faire courir l'IA à l'Humanité et les interrogations auxquelles elles aboutissent. Combien de fois n'avons-nous pas été confrontés à des questions soulevant tantôt les risques d'une prise de pouvoir par les machines (façon *Terminator*), tantôt d'une société dirigée par les algorithmes (façon *Westworld*), tantôt encore d'un monde dont la perception ne serait que le résultat d'une vaste simulation mise en place par des machines ou des ordinateurs (sur le modèle de *Matrix*). Il est regrettable que le public soit souvent encouragé à ne s'intéresser qu'aux scénarios préconçus de ces récits et non aux questionnements qu'ils pourraient soulever sur l'évolution non pas des machines, mais bien de notre propre humanité. Car, et c'est là sans doute une piste de réflexion à explorer, en plaçant l'homme face à la machine dans le contexte d'une approximation toujours plus précise entre les fonctions cognitives humaines et les fonctions de traitement de systèmes inertes, c'est notre humanité avant tout qui se doit d'être scrutée. En d'autres termes, l'intelligence artificielle en dit plus sur l'homme que sur la machine.

Pour des raisons fort variables, plusieurs nations sont résolument déterminées à intégrer l'IA au sein de leurs organisations militaires. Certaines puissances militaires et technologiques ont très tôt envisagé l'association de systèmes autonomes au cœur de leurs procédures décisionnelles. Cette exigence découlait, dans ce cas, des caractéristiques mêmes des armes qu'il s'agissait alors de mettre en œuvre : les missiles nucléaires. La crédibilité de la dissuasion supposait l'existence d'une capacité de seconde frappe dont l'actionnement devait être préservé en toutes circonstances, y compris dans l'hypothèse d'une destruction quasi totale des systèmes de défense et du territoire. En l'absence de tout commandement humain, un système autonome devait être en mesure de lancer une dernière salve contre des objectifs prédéfinis. D'autres nations encore se sont engagées sur la voie de l'IA en espérant accomplir un saut technologique qualitatif de plusieurs générations (sans passer par les générations intermédiaires) à même de renverser l'équilibre des forces découlant de la supériorité informationnelle détenue par certaines nations. C'est le cas de la Chine et, dans une certaine mesure, de la Russie. Quant aux autres nations, leur ralliement à l'IA à des fins militaires répond principalement à des exigences d'interopérabilité et de maintien d'un seuil technologique critique leur permettant de s'associer d'une manière ou d'une autre au dispositif de leur allié principal : les États-Unis, pour ce qui est des membres de l'OTAN.

La question de la domination de l'homme par la machine est un thème qui a connu de nombreux rebondissements parmi les plus éminents experts, en ce compris des scientifiques et technologues investis dans le domaine de l'intelligence artificielle. La perspective d'un contrôle par la machine du sort de l'Humanité s'appuie cependant sur un postulat de base erroné : l'idée selon laquelle la connaissance et la compétence permettraient à leur détenteur d'exercer une domination totale sur ceux qui ignorent. En d'autres termes, derrière le scénario d'une domination de l'Humanité par les machines intelligentes se cache le présupposé selon lequel plus un système démontre des aptitudes, des prouesses et un savoir élaboré, plus il serait tenté d'exercer une domination de plus en plus absolue. Jean-Gabriel Ganascia se montre catégorique à propos de ce mythe : « Jusqu'à présent, dans l'Histoire, il n'en est jamais allé ainsi. Les mathématiciens et les chercheurs n'ont jamais acquis un grand pouvoir du fait des connaissances qu'ils avaient acquises. De même, si les ingénieurs sont de temps à autres parvenus à bâtir des empires industriels, cela ne les a généralement pas portés au

pouvoir politique. On honore les personnes qui découvrent des connaissances neuves, on les rémunère lorsqu'elles maîtrisent des savoirs mieux que les autres, mais on n'abdique pas devant elles ». Dès lors, ajoute Ganascia, il y a peu de chance que des entités mues par des programmes d'intelligence artificielle puissent un jour nous dominer tout simplement parce qu'elles auraient acquis plus de compétences que les hommes. Cependant, le risque qu'un pouvoir politique élude ses responsabilités en déléguant certaines décisions à des machines derrière des critères de prétendue objectivité est très probable. Dans certains secteurs, comme l'éducation et la sélection des étudiants ayant accès aux études supérieures, des algorithmes sont d'ores et déjà au cœur de la prise de décision.

Les annonces fracassantes de l'imminence d'une transformation radicale et en profondeur de l'essence même de la guerre du fait de l'irruption d'une technologie nouvelle furent légions au cours de l'Histoire. Jamais un modèle de guerre, fût-il bâti sur des technologies aussi inédites qu'exotiques, n'a remplacé de manière définitive un ancien mode de confrontation. Les méthodes de guerre s'ajoutent aux précédentes plutôt qu'elles ne se substituent les unes aux autres. Il est très probable que les armements ou méthodes de combat qui découleront des technologies de l'intelligence artificielle ne dérogeront pas à cette règle. Loin des certitudes que semble offrir aux yeux de certains experts et observateurs l'avènement de l'IA dans le champ militaire, c'est vers un continent encore à découvrir que nous emmènent la science et la technologie de demain.

Liste des abréviations et acronymes

ACT	Allied Command Transformation
ADM	Arme de destruction massive
APL	Armée populaire de libération
BATX	Baidu, Alibaba, Tencent, Xiaomi
BIOS	Basic Input Output System
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance
CBO	Congressional Budget Office
CBRN	Chemical, Biological, Radiological, Nuclear
CCAC	Convention sur certaines armes conventionnelles
CFTC	Commodity Futures Trading Commission
CICDE	Centre interarmées de concepts, de doctrines et d'expérimentation
CICR	Comité international de la Croix-Rouge
CIFCM	Centre d'innovation et de fusion civilo-militaire
CNRS	Centre national pour la recherche scientifique
DARPA	Defense Advanced Research Project Agency
DIH	Droit international humanitaire
DIU	Defence Innovation Unit
DNN	Deep Neural Network
DOD	Department of Defense (Département de la Défense américain)
DSTL	Defence Science and Technology Laboratory
ERCS	Emergency Rocket Communications System
ESET	Enjoy Safer Technology
FAIR	Facebook Artificial Intelligence Research
FET	Future and emerging technology
FLOPS	Floating-point operations per second
FRA	Fondation (russe) pour la recherche avancée
GAFAMITIS	Google, Apple, Facebook, Amazon, Microsoft, IBM, Twitter, Intel et Salesforce
GAI	General artificial intelligence
GNR	Génétique, nanotechnologie, robotique
GPS	Global Positioning System
GPU	Graphical processor unit
HALE	Haute altitude, longue endurance

HFT	High frequency trading
IA	Intelligence artificielle
IAG	Intelligence artificielle générale
ICBM	Intercontinental ballistic missile
ICML	International Conference on Machine Learning
IEET	Institute for Ethics and Emerging Technologies
IFRI	Institut français des relations internationales
IISS	International Institute for Strategic Studies
IM	Information management
IO	Information operations
IS	Information superiority
ISR	Intelligence, surveillance and reconnaissance
JAIC	Joint Artificial Intelligence Center
LAWS	Lethal autonomous weapon system
LRSO	Long-range stand-off
MAD	Mutual assured destruction
MALE	Moyenne altitude longue endurance
HBSP	Human Brain Simulation Project
MIRI	Machine Intelligence Research Institute
MIT	Massachusetts Institute of Technology
NATU	Netflix, Airbnb, Tesla and Uber
NBC	Nucléaire, biologique, chimique
NBIC	Nanotechnology, biotechnology, information technology and cognitive science
NC3	Nuclear command, control and communications
NCIA	NATO Communication and Information Agency
NNI	National Nanotechnology Initiative
NORAD	North American Air Defense Command
ONG	Organisation non-gouvernementale
ONR	Office of Naval Research
OODA	Observation, orientation, décision, action
R&D	Recherche & développement (Research & Development)
RMA	Revolution in military affairs
SAGE	Semi-Automatic Ground Environment
SALA	Système d'armes létal autonome

SAPE	Survivable Adaptive Planning Experiment
SAPIENT	Sensing for Asset Protection using Integrated Electronic Networked Technology
SEC	Securities and Exchange Commission
SIPRI	Stockholm International Peace Research Institute
SLBM	Submarine-launched ballistic missile
STO	Science & Technology Organisation
TCI	Technologies des communications et de l'information
TOS	Third Offset Strategy
TRACE	Target Recognition and Adaptation in Contested Environments
UEFI	Unified Extensible Firmware Interface
UHF	Ultra-haute fréquence
URSS	Union des républiques socialistes soviétiques
USAF	United States Air Force
XAI	eXplainable Artificial Intelligence

Bibliographie

1. Ouvrages

ALEXANDRE, Laurent, *La guerre des intelligences : comment l'intelligence artificielle va révolutionner l'éducation*, Paris, Edition Jean-Claude Lattès, 2017.

ALLEN, Greg, CHAN, Taniel, *Artificial Intelligence and National Security*, Belfer Center for Science and International Affairs, 2017.

BENASAYAG, Miguel, *La tyrannie des algorithmes. Conversation avec Régis Meyran,* Paris, Textuel, coll. Conversations pour demain, 2019.

BESNIER, Jean-Michel, *Demain, les post-humains. Le futur a-t-il encore besoin de nous ?*, Paris, Fayard, 2020.

BILAL, Enki, et al., *Intelligence artificielle. Enquête sur ces technologies qui changent nos vies*, Paris, Flammarion, Champs/Actuel, 2018.

BOULANIN, Vincent, (eds), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk, Volume I: Euro-Atlantic Perspectives*, Solna, Stockholm International Peace Research Institute (SIPRI), mai 2019.

BRETON, Philippe, *Une histoire de l'informatique*, Paris, Seuil, 1990, p. 129.

CREVIER, Michel, *À la recherche de l'intelligence artificielle*, Paris, Flammarion, coll. Champs, 1997.

GANASCIA, Jean-Gabriel, *Le mythe de la Singularité : faut-il craindre l'intelligence artificielle ?*, Paris, Éditions du Seuil, Points/Essais, 2017.

HOROWITZ Michael C. & al., *Strategic Competition in an Era of Artificial Intelligence*, Center for a New American Security, 2018.

JOUSSET-COUTURIER, Béatrice, *Le transhumanisme : faut-il avoir peur de l'avenir ?*, Paris, Éditions Eyrolles, 2016.

KENNEDY, Paul, *Naissance et déclin des grandes puissances : transformations économiques et conflits militaires entre 1500 et 2000*, traduction de Marie-Aude Cochez et Jean-Louis Le brave, Paris, Payot, coll. Petite bibliothèque Payot, 1991.

LAFONTAINE, Céline, *L'Empire cybernétique : des machines à penser à la pensée machine*, Paris, Éditions du Seuil, Essai, 2004.

LE BORGNE, Claude, *La Guerre est morte, mais on ne le sait pas encore*, Grasset, 1987.

LEE, Kai-Fu, *AI Superpowers: China, Silicon Valley, and the New World Order*, Boston/New York, Houghton Mifflin Harcourt, 2018.

MORANGE, Michel, *Histoire de la biologie moléculaire*, Paris, La Découverte, 1994.

MORAVEC, Hans, *The Role of Raw Power in Intelligence,* Stanford, Stanford Libraries, 12 mai 1976.

NILSON, N. J., *The Quest for Artificial Intelligence: A History of Ideas and Achievements?,* Cambridge, Cambridge University Press, 2009.

2. Rapports

Commission européenne, *A Definition of AI: Main Capabilities and Disciplines*, Definition developed for the purpose of the AI High-Level Expert Group's deliverables, Bruxelles, Commission européenne, 8 avril 2019.

CRS (service de recherche du Congrès américain), « U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress », CRS, 2018.

DYÈVRE, Axel, GOETZ, Pierre, FERRANDO, Florence, *Intelligence artificielle : applications et enjeux pour les Armées*, Paris, Compagnie européenne d'intelligence stratégique (CEIS), coll. Les Notes Stratégiques, septembre 2018.

GEIST, Edward, LOHN, Andrew J., (eds), *How Might Artificial Intelligence Affect the Risk of Nuclear War?*, RAND Corporation, Santa Monica (Calif.), Serie Perspective, 2018.

HOADLEY, Daniel S., Lucas, NATHAN J., *Artificial Intelligence and National Security*, CRS, 2018.

KOSTOPOULOS, Lydia, *The Role of Data in Algorithmic Decision-Making: A Primer*, United Nations Institute for Disarmement Research (UNIDIR), 2019, http://www.unidir.org.

NOCETTI, Julien, *Intelligence artificielle et politique internationale : les impacts d'une rupture technologique*, Paris, Institut français des relations internationales, Études de l'IFRI, novembre 2019.

O'HUNDLEY, Richard, *Past Revolutions, Future Transformations: What Can the History of Revolutions in Military Affairs Tell Us About Transforming the U.S. Military?*, Santa Monica (Calif.), RAND Corporation, 1999.

SAYLER, Kelly M., *Artificial Intelligence and National Security*, CRS Report, R45178, 21 novembre 2019.

SÉGAL, Jérôme, *Le zéro et le un. Histoire de la notion scientifique d'information au 20ème siècle*, Paris, Syllepses, coll. Matériologiques, 2003.

SHEPPARD, Lindsey R., & al., *Artificial Intelligence and National Security: the Importance of the Ecosystem*, CSIS, novembre 2018.

THIBODEAUX, Maxwell, KAPLAN, Richard, SMITH, Anthony, CLEMA, Joe K., « A Framework for Understanding the IO: C4ISR Relationship », Colloque Command and Control Research Program, papers/062, 2006.

3. Articles

ALEXANDRE, Frédéric, TISSERON, Serge, « Où sont les vrais dangers de l'intelligence artificielle ? », *Pour la Science*, Dossier numéro 87, avril-juin 2015.

ALLEN, G., KANIA, E. (8 septembre 2017). « China is using America's own plan to dominate the future of artificial intelligence », Foreign Policy, retrieved from https://foreignpolicy.com/2017/09/08/china-is-using-americas-own-plan-todominate- the-future-of-artificial-intelligence/

ANDERSON, Wendy R., TOWNSEND, Jim, « As AI Begins to Reshape Defense, Here's How Europe Can Keep Up », *Defense One*, 18 mai 2018.

BENGIO, Yoshua, « La révolution de l'apprentissage profond », *Pour la Science Hors-Série*, Big Data, numéro 98, février 2018.

BESNIER, Jean-Michel, « Les nouvelles technologies vont-elles réinventer l'homme ? », *Études*, 2011, volume 6, tome 414, pp. 736-772.

BEZOMBES, Patrick, « Intelligence artificielle et robots militaires », *Défense & Sécurité Internationale Hors-Série*, numéro 65, avril-mai 2019.

BLANK, Stephen, « Can Information Warfare Be Deterred ? », *Defense Analysis*, Vol. 17, No. 2, 2001.

BREIMAN, L. « Statistical Modeling: The Two Cultures », *Statistical Science*, Vol. 16, No. 3, 2001.

BRYON-PORTET, Céline, « Du devoir de soumission au devoir de désobéissance ? Le dilemme militaire », *Res Militaris*, *cf.* http://resmilitaris.net/ressources/10123/66/5_res_militaris_article_bryon-portet_texte_inte_gral.pdf

CARR, Nicholas, « It's Not a Bug, it's a Feature. Trite – or Just Right? », *Wired*, *cf.* https://www.wired.com/story/its-not-a-bug-its-a-feature

DEBLOCK, Christian, « Introduction : innovation et développement chez Schumpeter », *Revue interventions économiques*, volume 46, 2012.

DONT, Barthélémy, « Amazon a dû se débarrasser d'une intelligence artificielle sexiste », *Slate*, 10 octobre 2018, *cf.* http://www.slate.fr/story/168413/amazon-abandonne-intelligence-artificielle-sexiste

DURHIN, Nathalie, « Systèmes d'armes létaux autonomes : ne pas mélanger juridique et éthique », *Les Cahiers de la Revue Défense Nationale*, 2019.

FLORIDI, Luciano, COWLS, Josh, « A Unified Framework of Five Principles for AI in Society », *Harvard Data Science Review*, juillet 2019, *cf.* https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/6

FRÉGNAC, Yves, Gilles LAURENT, « Where is the Brain in the Human Brain Project? », *Nature*, Vol. 513, 4 septembre 2014, pp. 27-29.

GESNY, Olivier, « Capter l'IA de demain au regard des enjeux de cyberdéfense », *Revue Défense Nationale*, dossier « L'intelligence artificielle et ses enjeux pour la Défense », mai 2019.

GOUBET, Fabien, « Une nouvelle crise secoue le Human Brain Project », *Le Temps*, 21 août 2018, *cf.* https://www.letemps.ch/sciences/une-nouvelle-crise-secoue-human-brain-project

HAUTECOUVERTURE, Benjamin, « Applications nucléaires de l'automatisation : rappels historiques », *Bulletin mensuel de l'Observatoire de la dissuasion*, sous la direction de MAÎTRE, Emmanuel, et TERTRAIS, Bruno, Paris, Fondation pour la recherche stratégique (FRS) & Direction générale des relations internationales et de la stratégie (DGRIS), numéro 67, été 2019.

HENNINGER, Laurent, « Espaces fluides et espaces solides : nouvelle réalité stratégique », *Revue Défense nationale*, octobre 2012, numéro 753.

HENROTIN, Joseph, « Le drone : figure aérienne du mal ? », *Défense & Sécurité Internationale Hors-Série*, numéro 30, juin – juillet 2013.

HOROWITZ, M. C. (15 mai 2018). « Artificial intelligence, international competition, and the balance of power", Texas National Security Review, retrieved from https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-thebalance- of-power/

IISS, « Big Data, Artificial Intelligence and Defence", *The Military Balance 2018*, IISS, 2018.

JEANGÈNE-VILMER, Jean-Baptiste, « Légalité et légitimité des drones armés », *Politique étrangère*, 2013/3.

JOHNSON, James, « Artificial Intelligence & Future Warfare: Implications for International Security », *Defense & Security Analysis*, Vol. 35, No. 2, 2019.

JOY, Bill, « Why The Future Doesn't Need Us », *Wired*, *cf.* https://www.wired.com/2000/04/joy-2/

KANIA, Elsa, « Chinese Sub Commanders May Get AI Help for Decision-Making », *Defense One*, 12 février 2018, <https://www.defenseone.com/ideas/2018/02/chinese-sub-commanders-may-get-ai-helpdecision-making/145906/?oref=d-river>

KAYSER-BRIL, Nicolas, « 'Explainable AI' Doesn't Work for Online Services – Now There's Proof », 12 novembre 2019, *cf.* https://algorithmwatch.org/en/story/explainable-ai-doesnt-work-for-online-services-now-theres-proof

KEMPF, Olivier, « IA, explicabilité et défense », *Revue Défense Nationale,* mai 2019.

KLARE, Michael T., « Skynet Revisited: The Dangerous Allure of Nuclear Command Automation », *Arms Control Today*, Arms Control Association, avril 2020, *cf*. https://www.armscontrol.org/act/2020-04/features/skynet-revisited-dangerous-allure-nuclear-command-automation

LE CUNN, Yann, *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond,* Paris, Odile Jacob, 2019.

LE MERRER, Erwan, TRÉDAN, Gilles, « The Bouncer Problem: Challenges to Remote Explainability », 3 octobre 2019, *cf.* https://arxiv.org/pdf/1910.01432v1.pdf

LOWTHER, Adam, MCGIFFIN, Curtis, « America needs a 'Dead hand' », *War on the Rocks*, 16 août 2019, *cf*. https://warontherocks.com/2019/08/america-needs-a-dead-hand/

LYE, Harry, « Skyborg: The US Air Force's Future AI Fleet", *Air Force Technology*, 28 août 2019, https://www.airforce-technology.com/features/skyborg-the-us-air-forces-future-ai-fleet

LYE, Harry, « RAF to Launch Swarming Drones in April », http://www.airforce-technology.com, 13 janvier 2020, *cf.* https://www.airforce-technology.com/news/raf-swarming-drones

MARCUM R. A., DAVIS C. H., SCOTT G. J., NIRVIN T. W., « Rapid Broad Area Search and Detection of Chinese Surface-to-Air Missile Sites Using Deep Convolutional Neural Networks », *Journal of Applied remote Sensing*, vol. 11, no. 4, octobre-décembre 2017.

MARKRAMN, Henry, MEIER, Karlheinz, LIPPERT, Thomas, GRILLNER, Sten, FRACKOWIAK, Richard, DEHAENE, Stanislas, KNOLL, Alois, SOMPOLINSKY, Haim, VERSTREKEN, Kris, DEFELIPE, Javier, GRANT, Seth, CHANGEUX, Jean-Pierre, SARIA, Alois, « Introducing the Human Brain Project », *Procedia Computer Science*, No. 7, 2011.

MIAILHE, Nicolas, « Géopolitique de l'Intelligence artificielle : le retour des empires ? », *Politique étrangère*, volume 3, 2018.

NEIRYNCK, Jacques, « L'ordinateur ne peut simuler le cerveau », *Le Temps,* 13 janvier 2019, *cf.* https://www.letemps.ch/opinions/lordinateur-ne-simuler-cerveau

NOCETTI, Julien, « La Chine, superpuissance numérique ? », dans Thierry DE MONTBRIAL, Dominique DAVID, *RAMSES 2019. Un monde sans boussole ?*, Paris, Institut français des relations internationales /Dunod, 2018.

NOCETTI, Julien, *Intelligence artificielle et politique internationale*, Paris, Institut français des relations internationales, Études de l'IFRI, novembre 2019.

PISTONO, Federico, YAMPOLSKIY, Roman V., « Unethical Research: How To Create a Malevolent Artificial Intelligence », *cf.* https://arxiv.org/ftp/arxiv/papers/1605/1605.02817.pdf

PORTNOFF, André-Yves, SOUPIZET, Jean-François, « Intelligence artificielle : opportunités et risques », *Futuribles*, volume 5, numéro 426, 2018.

RAMAMOORTHY, Anand, YAMPOLSKIY, Roman, « Beyond MAD ? The Race for Artificial General Intelligence », *ITU Journal: ICT Dicoveries*, Special Issue No. 1, 2 février 2018.

ROHMER, Jean, « Le transmachinisme : et si les machines évoluaient indépendamment de l'homme ? », *The Conversation*, *cf*. https://theconversation.com/le-transmachinisme-et-si-les-machines-evoluaient-independamment-de-lhomme-138367

SAALMAN, Lora, « Fear of False Negatives : AI and China's Nuclear Posture », *Bulletin of Atomic Scientists*, 24 avril 2018, *cf*. https://thebulletin.org/2018/04/fear-of-false-negatives-ai-and-chinas-nuclear-posture/

TUCKER, P., « How AI Will Transform Anti-Submarine Warfare », *Defense One*, 1er juillet 2019.

VERHAEGEN, Jacques, « Le refus d'obéissance aux ordres manifestement criminels: pour une procédure accessible aux subordonnés », *Revue internationale de la Croix-Rouge*, volume 84, numéro 845, mars 2002.

WIENER, Norbert, « Man and the Machine », (interview with Norbert Wiener), *Challenge: The Magazine of Economic Affairs*, No. 7, 1959.

XIANG, Li, « Artificial Intelligence and Its Impact on Weaponization and Arms Control », in Lora SAALMAN (ed.), *The Impact of Artificial Intelligence on Strategic Stability and Nuclear Risk*, Solna (Sweden), Stockholm International Peace Research Institute (SIPRI), octobre 2019.

YAMPOLSKIY, Roman V., « Predicting Future AI Failures from Historic Examples », *Foresight*, novembre 2018, https://www.researchgate.net/publication/329225671_Predicting_future_AI_failures_from_historic_examples

YAMPOLSKIY, Roman V., « Unexplainability and Incomprehensibility of Artificial Intelligence », *cf*. https://arxiv.org/ftp/arxiv/papers/1907/1907.03869.pdf

4. Sites Internet

ARKIN, Ronald C., « The Case for Ethical Autonomy in Unmanned Systems », *cf.* https://www.cc.gatech.edu/ai/robot-lab/online-publications/Arkin_ethical_autonomous_systems_final.pdf

FRANKE, Ulrike Esther, *Flash Wars: Where Could an Autonomous Weapons Revolution Lead Us ?*, https://www.ecfr.eu/article/Flash_Wars_Where_could_an_autonomous_weapons_revolution_lead_us

INSINNA, Valerie, « Introducing Skyborg, Your New AI Wingman », C4ISRNET, 14 mars 2019.

KELLER, John, « DARPA TRACE Program Using Advanced Algorithms, Embedded Computing for Radar Target Recognition », *Military Aerospace & Electronics*, 24 juillet 2015, *cf*. https://www.militaryaerospace.com/computers/article/16714226/darpa-trace-program-using-advanced-algorithms-embedded-computing-for-radar-target-recognition

À propos du CESD et de ses publications

L'Institut royal supérieur de défense (IRSD) est le centre de réflexion de référence spécialisé du ministère de la Défense belge. Il est notamment chargé, tant au niveau national qu'international, de mener la recherche interdisciplinaire dans le domaine de la politique de sécurité et de défense au service de la société.

Au sein de l'IRSD, le Centre d'études de sécurité et défense (CESD) est chargé de nourrir la réflexion présidant à l'élaboration des politiques futures dans le domaine de la sécurité et de la défense grâce aux résultats de ses recherches. À cet effet, le CESD délivre des analyses objectives et développe des visions à plus long terme afin d'optimiser la réflexion politique et d'attirer l'attention des décideurs sur des points cruciaux dans le domaine de la politique de sécurité et de défense.

La recherche au sein du CESD est consacrée à l'étude des tendances politiques, militaires, institutionnelles, technologiques, socio-économiques et idéologiques qui sont susceptibles d'avoir un impact sur la naissance, le développement et les conséquences des crises et des conflits à travers le monde.

Le CESD publie le résultat de ses recherches dans trois séries, accessibles sur son site Internet et/ou en format papier :

- la revue *Sécurité & Stratégie* publie les études en matière de sécurité et de défense prévues par le programme annuel de recherche scientifique de la Défense belge. Les études proposées sont basées sur les lignes de recherche, sur l'évolution de l'environnement sécuritaire, sur les dispositions de « La vision stratégique pour la Défense » du gouvernement belge, sur les orientations données par l'état-major et sur l'expertise des chercheurs ;
- la série *Focus Papers* publie les résultats des recherches qui n'entrent pas dans le cadre du programme annuel de recherche scientifique de la Défense proprement dit. Les études proposées concernent des problématiques *ad hoc* ainsi que les résultats de recherche des stagiaires du CESD ;
- les *e-Notes* sont des articles traitant de sujets liés à l'actualité.

Les organisations de défense face aux défis de l'intelligence artificielle

Généralement associées à l'imaginaire issu de la science-fiction, littéraire ou cinématographique, utopique ou dystopique, les technologies de l'intelligence artificielle ont investi ces vingt dernières années des domaines de plus en plus vastes de la vie des individus et des collectivités. Il n'est pas un secteur qui, aujourd'hui, ne soit influencé, affecté, normé ou administré par l'IA et ses algorithmes avancés. Les forces armées, à l'image des sociétés qu'elles servent, n'échappent pas à cette tendance. Désormais, les technologies de l'intelligence artificielle, appuyées par des algorithmes complexes et la multitude des données provenant de l'emploi fait des réseaux numériques les plus diversifiés, ont étendu leur emprise – et peut-être même leur empire – sur l'ensemble des aspects de la vie humaine. La guerre en fait partie et certains s'interrogent déjà savoir si elle persistera à demeurer une dialectique de volontés humaines.

L'intelligence artificielle, qui se situe d'ores et déjà au cœur d'une course technologique réelle, s'apprête à bouleverser tant les processus de décision politique que les modes d'action militaire des forces armées. Comme chaque révolution technique, celle de l'IA participera à l'émergence de nouveaux acteurs et la relégation d'anciens, quand il ne s'agira pas purement et simplement de leur effacement. Qu'elles en maîtrisent les applications ou qu'elles en subissent les effets, les organisations de défense de demain connaîtront des transformations sans précédent. Entre « Singularité », « hyper-guerres », « guerre algorithmiques », « systèmes d'armes autonomes », « cryptographie quantique », « flash wars » et « IApocalyspes », les concepts, les perspectives et les scénarios s'entrechoquent et laissent entrevoir un futur incertain. Bienvenue dans l'ère stratégique de l'intelligence artificielle.

Alain De Neve, *politologue de formation, est chercheur Développements technologiques en matière de défense au Centre d'études de sécurité et défense de l'Institut royal supérieur de défense.*

Source photo : https://pixabay.com/fr/illustrations/femme-circuits-5899198/, Image par Gerd Altmann de Pixabay
Source photo : https://www.belgium-naval-and-robotics.be/fr/page/3/
Source photo : https://www.af.mil/News/Photos/igphoto/2000694430/, U.S. Air Force photo/Staff Sgt. Brian Ferguson
Source photo : https://twitter.com/jmscaronte/status/1283135825836863488, photo de Joaquin Aldecoa

Institut royal supérieur de défense
Avenue de la Renaissance 30
1000 Bruxelles - Belgique
www.defence-institute.be

ISSN 0770-9005

LA DÉFENSE