

NOVEMBRE 2020

Observatoire de la politique de défense américaine

Positionnement des acteurs américains sur les questions éthiques liées à l'intelligence artificielle de défense

FREDERIC COSTE

AVEC LA PARTICIPATION DE NICOLAS MAZZUCCHI

FONDATION
pour la **RECHERCHE**
STRATÉGIQUE

Marché n° 2018 1050 118 198

EJ court 180 005 17 10

notifié le 18 octobre 2018

réunion de lancement :

14 novembre 2018

avec le soutien de la



Avertissement

Les propos exprimés dans ce rapport n'engagent que leurs auteurs et ne constituent en aucune manière une position officielle de la Fondation pour la Recherche Stratégique ou du ministère des Armées.

SOMMAIRE

LISTE DES ABRÉVIATIONS

RÉSUMÉ.....	1
INTRODUCTION	4
EFFORTS DE RÉFLEXION SUR LES QUESTIONS ÉTHIQUES AU SEIN DE L'ÉCOSYSTÈME AMÉRICAIN	7
1. CADRE INSTITUTIONNEL GÉNÉRAL.....	7
1.1. Construction progressive du cadre institutionnel	7
1.1.1. 2016 et la pose des premières pierres de l'édifice	7
1.1.2. Renforcement des structures de coordination de haut niveau (2018-2019)	8
1.2. Publication de la Stratégie nationale en IA	9
2. CADRE SPÉCIFIQUE À LA DÉFENSE	12
2.1. Dispositions du National Defense Authorization Act (NDAA) 2019	13
2.1.1. Nomination d'un haut responsable en charge de la coordination des activités de développement et démonstration des systèmes recourant à l'IA.....	13
2.1.2. Création d'une Commission de sécurité nationale sur l'Intelligence artificielle (« AI Commission »).....	13
2.1.3. Renforcement des fonctions du Joint Artificial Intelligence Center (JAIC)	14
2.2. Mission confiée au Defense Innovation Board (DIB)	15
2.3. Stratégie en IA du DoD	15
2.4. Prolongements de 2019.....	18
2.5. Bilan.....	18
3. EFFORTS POUR TRAITER LES THÉMATIQUES ÉTHIQUES LIÉES AUX APPLICATIONS IA DE DÉFENSE	19
3.1. Principaux documents dédiés spécifiquement aux dimensions éthiques.....	19
3.1.1. Directive 3000.09	19
3.1.2. Principes éthiques du DoD.....	21
A. Contenu du document	21
B. Limites des Principes.....	23

3.2.	Principales structures impliquées dans le traitement des dimensions éthiques de l'IA de défense	24
3.2.1.	Le Defense Innovation Board.....	24
3.2.2.	L'AI Commission	25
3.2.3.	Le JAIC.....	26
4.	THÉMATIQUES SOUS-TENDANT LE TRAITEMENT DES QUESTIONS ÉTHIQUES	28
STRUCTURES INTERNATIONALES TRAITANT DES QUESTIONS ÉTHIQUES		33
1.	MOUVEMENTS DE CONTESTATION DES SALA	33
1.1.	Principales critiques à l'égard des SALA	34
1.2.	Modes d'action et demandes	35
2.	PRINCIPALES ORGANISATIONS INTERNATIONALES PRENANT EN CHARGE DES DÉBATS SUR LES PROBLÉMATIQUES ÉTHIQUES LIÉES À L'IA	37
2.1.	Nations Unies	37
2.1.1.	UNESCO	38
2.1.2.	Union internationale des télécommunications (UIT)	38
2.1.3.	UN Convention on Certain Conventional Weapons (CCW)	39
2.2.	Union européenne	42
2.2.1.	Documents majeurs	43
A.	Communication de la Commission européenne (avril 2018)	43
B.	Déclaration commune de soutien à l'approche de l'UE (avril 2018)	44
C.	Publication des Ethics Guidelines for Trustworthy Artificial Intelligence (avril 2019)	45
2.2.2.	Discussion autour des SALA.....	47
2.3.	Organisation de coopération et de développement économiques (OCDE)	49
2.4.	G7	50
2.5.	Global Partnership on Artificial Intelligence (GPAI)	51
3.	TERMES DES DÉBATS	52
3.1.	Thématiques faisant l'objet d'un traitement	52
3.2.	Considérations juridiques	55
3.3.	Réflexions sur les outils à mettre en place	56
4.	TRAVAUX AU SEIN DE L'OTAN	57
5.	EFFORTS DE STANDARDISATION AU NIVEAU INTERNATIONAL	59
5.1.	Prise en charge par l'International Organization for Standardization (ISO)	59
5.2.	Traitement dans le cadre de l'Institute of Electrical and Electronics Engineers (IEEE)	62

INSERTION DES ACTEURS AMÉRICAINS DANS LES STRUCTURES INTERNATIONALES 65

1. POSITIONNEMENT DANS LE CADRE DES TRAVAUX DE LA CONVENTION SUR CERTAINES ARMES CONVENTIONNELLES	66
2. UNE RECHERCHE DE PARTENAIRES PRIVILÉGIÉS.....	68
2.1. Relations bilatérales	68
2.1.1. « Relation spéciale » et IA	68
2.1.2. Partenariat avec Singapour.....	70
2.2. Création de partenariats multilatéraux dédiés	71
2.3. Un effort de mobilisation dans le cadre de l'OTAN.....	72
3. DES CADRES MULTILATÉRAUX INVESTIS AVEC PRUDENCE.....	72
4. DOMINER L'ÉDIFICATION DES STANDARDS TECHNIQUES INTERNATIONAUX.....	74
4.1. Efforts chinois en matière de standardisation	74
4.2. Efforts des États-Unis.....	77

ANNEXE 1

ACTIVITÉS DE LA NATO-STO LANCÉES DANS LE CADRE DE LA THÉMATIQUE « INTELLIGENCE ARTIFICIELLE ET BIG DATA POUR LA PRISE DE DÉCISION MILITAIRE »	79
--	-----------

ANNEXE 2

ACTIVITÉS DE LA NATO-STO LANCÉES DANS LE CADRE DE LA THÉMATIQUE « AUTONOMIE »..	81
--	-----------

Liste des abréviations

ACT	<i>Allied Command Transformation</i>	NIAG	<i>NATO Industrial Advisory Group</i>
AFRL	<i>Air Force Research Laboratory</i>	NDAA	<i>National Defense Authorization Act</i>
AHEG	<i>Ad Hoc Expert Group</i>	NDS	<i>National Defense Strategy</i>
AIM	<i>Augmenting Intelligence using Machines</i>	NIAG	<i>NATO Industrial Advisory Group</i>
ANSI	<i>American National Standards Institute</i>	NIST	<i>National Institute for Standards and Technology</i>
BATX	Baidu, Alibaba, Tencent, Xiaomi	NITRD	<i>Networking and Information Technology Research and Development Committee</i>
CCW	Certain Conventional Weapons	NSF	<i>National Science Foundation</i>
CESI	<i>China Electronics Standardization Institute</i>	NSTC	<i>National Science and Technology Council</i>
COMEST	Commission mondiale d'éthique des connaissances scientifiques et des technologies	NCI	<i>NATO Communications and Information Agency</i>
DARPA	<i>Defense Advanced Research Projects Agency</i>	OCDE	Organisation de coopération et de développement économiques
DIB	<i>Defense Innovation Board</i>	OODA	<i>observe–orient–decide–act</i>
DoD	<i>Department of Defense</i>	OSS	<i>Open Source Software</i>
ECPAIS	<i>Ethics Certification Program for Autonomous and Intelligent Systems</i>	OTAN	Organisation du Traité de l'Atlantique Nord
EPSRC	<i>Engineering and Physical Sciences Research Council</i>	PMIA	Partenariat mondial sur l'intelligence artificielle
ESG	<i>Executive steering group</i>	RTG	<i>Research Task Group</i>
G2IA	Groupe international d'experts sur l'intelligence artificielle	OTAN	Organisation du Traité de l'Atlantique Nord
GGE	<i>Group of Governmental Experts</i>	PMIA	Partenariat mondial sur l'intelligence artificielle
GPAI	<i>Global Partnership on Artificial</i>	RTG	<i>Research Task Group</i>
IA	Intelligence artificielle	SALA	Systèmes d'armes létaux autonomes
ICRAC	<i>International Committee for Robots Arms Control</i>	S&T	Science et technologie
IEC	<i>International Electrotechnical Commission</i>	STB	<i>NATO Science and Technology Board</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>	STEM	<i>Science, Technology, Engineering, and Mathematics</i>
IJCAI	<i>International Joint Conferences on Artificial Intelligence</i>	STO	<i>NATO Science and Technology Organization</i>
ISO	<i>International Organization for Standardization</i>	UE	Union Européenne
JAIC	<i>Joint Artificial Intelligence Center</i>	UNESCO	<i>United Nations Educational, Scientific and Cultural Organization</i>
MLAI	<i>Machine Learning and AI Committee</i>	UNIDIR	<i>United Nations Institute for Disarmament Research</i>
MURI	<i>Multidisciplinary University Initiative</i>	UIT	Union International des Télécommunications
NATO	<i>North Atlantic Treaty Organization</i>	UNESCO	<i>United Nations Educational, Scientific and Cultural Organization</i>
		UNIDIR	<i>United Nations Institute for Disarmament Research</i>

Positionnement des acteurs américains sur les questions éthiques liées à l'intelligence artificielle de défense

Résumé

Dès la publication des premiers documents généraux d'orientation par l'Exécutif américain (2016), les questions éthiques que soulèvent les utilisations de l'intelligence artificielle (IA) ont été abordées. Leur étude a été intégrée aux politiques de soutien conçues par la Maison blanche. Les parlementaires ont accompagné cet effort, tout comme certains États fédérés. Surtout, au sein de l'écosystème de l'IA américain, de nombreux acteurs, très différents (fédérations professionnelles, grands groupes industriels, *think tanks*...), ont décidé de produire leurs propres analyses ou de participer à des initiatives collectives. Cette « effervescence » a notamment abouti à la publication de nombreux textes, en particulier des codes, principes et lignes directrices.

La prise en compte des questions éthiques liées au développement de l'IA concerne aussi les applications en matière de défense. Elle a commencé relativement tôt. Dès 2012, le DoD a posé un cadre particulier pour la conception et l'emploi des systèmes d'armes létaux autonomes (SALA), la *Directive 3000.09*. Cette dernière a permis de présenter les définitions que le Département retient pour les différents types d'autonomie et de systèmes armés autonomes, mais également de préciser les règles particulières devant s'appliquer à la création et aux utilisations de ces dispositifs. Les réflexions ont ensuite été élargies, dépassant la seule notion d'autonomie. L'une des 5 composantes de la *Department of Defense Artificial Intelligence Strategy*, adoptée fin 2018, est ainsi de produire les efforts nécessaires à la réflexion en matière éthique et sur les questions de sécurité d'emploi. Ce travail a abouti à la publication d'un jeu de principes éthiques en février 2020. La réflexion sur les outils pour mettre en œuvre ces principes a été confiée au *Joint Artificial Intelligence Center* (JAIC), qui a engagé plusieurs chantiers.

Dans une certaine mesure, la position américaine vis-à-vis de l'IA de défense est caractérisée par des ambitions contradictoires. L'objectif d'intégrer rapidement l'IA au sein du DoD est affirmé. Il s'agit d'accélérer un mouvement engagé : certaines technologies étant assez matures, il faut le plus rapidement possible achever le développement d'applications, les tester et les intégrer aux systèmes et organisations existants (et plus seulement renforcer la R&D). La nécessité de traiter les questions éthiques est en parallèle énoncée. Or, si l'identification des interrogations éthiques que soulève l'IA de défense est déjà un travail complexe, la mise au point de réponses l'est encore plus, demandant du temps et des itérations.

En dehors de la publication des principes, garde-fous communs à tous les services concernés par l'IA, plusieurs réponses sont apportées aux problèmes éthiques. Dans les documents institutionnels, apparaît tout d'abord l'idée que l'Intelligence artificielle doit être « au service des personnels », devant permettre de les autonomiser et non de les remplacer. Elle doit les décharger des tâches fastidieuses. De même, est affirmé le besoin de rendre les dispositifs compréhensibles par les opérateurs. Enfin, ces mêmes documents précisent que les personnels, notamment militaires, du DoD doivent être acteurs des débats sur les questions éthiques.

L'étude des questions juridiques a par ailleurs été érigée en priorité institutionnelle. Les textes de référence du Département indiquent que leur traitement doit impliquer la conformité à des ensembles de droit, nationaux et internationaux, et même l'éventuelle création de nouvelles références juridiques.

Dans les textes institutionnels, un lien est clairement fait entre questions éthiques et techniques de test et d'évaluation des systèmes dotés d'IA. Il est affirmé qu'une partie des problèmes éthiques peut être gérée grâce à une parfaite connaissance des domaines d'emploi des systèmes et si leur prévisibilité est assurée. Disposer de systèmes robustes, fiables et sécurisés est en soi une réponse à nombre de questions éthiques, qui justifie les efforts pour se doter de méthodes et outils de test, évaluation, vérification et certification. Les autorités militaires américaines ont répété que le pays devait être *leader* mondial dans ce domaine.

Enfin, le traitement des questions éthiques est très clairement mis en rapport avec la protection des valeurs (notamment démocratiques) du pays.

Au niveau international, de nombreuses enceintes ont décidé de mettre à leurs agendas le traitement des problèmes éthiques liés au développement de l'IA (UNESCO, OCDE, G7, G20, Union européenne...). Quelques-unes ont même traité spécifiquement des applications militaires (en particulier de la thématique des SALA). Au sein de l'OTAN, de nombreuses actions ont également été lancées. Les États-Unis ont investi la plupart des initiatives internationales qui servent à créer des cadres globaux pour le développement de l'IA. Ils cherchent également, pour les applications militaires, à mettre en place des partenariats, bilatéraux et multilatéraux, avec certains de leurs alliés et partenaires. Ces partenariats, qui concernent souvent des actions communes de R&D, permettent aux représentants américains de présenter leurs réflexions et textes en matière éthique.

Au sein des organisations internationales, les délégations américaines réaffirment leur volonté d'une IA respectueuse des valeurs démocratiques. La dimension axiologique du discours sert à polariser les positionnements. La montée en puissance chinoise en matière d'IA est décrite comme dangereuse du fait des applications militaires investiguées et du renforcement de la compétitivité économique du pays. Plus fondamentalement, les dirigeants américains dénoncent le fait que Pékin veut utiliser l'Intelligence artificielle pour asseoir son régime autoritaire, en particulier pour assurer un contrôle social et géographique des populations et réprimer les mouvements de contestation. Dans les documents institutionnels et les positions officielles, il est affirmé que les États-Unis doivent mobiliser leurs alliés et partenaires pour parvenir à créer des cadres internationaux pour l'IA protecteurs des valeurs démocratiques.

La vivacité des efforts généraux de réflexion dans le pays, la pression créée par la campagne internationale « *Stop Robots Killer* », le développement des discussions sur les SALA au sein

de la *UN Convention on Certain Conventional Weapons* (CCW), les efforts de montée en puissance de la Chine dans le domaine de l'IA et la volonté de rassurer en interne les personnels du DoD constituent, entre autres facteurs, de puissants éléments favorisant l'organisation de réflexions sur les questions éthiques par le Département ou la participation à celles qui se sont développées au niveau international. Un sentiment d'urgence face à un risque de « déclassement stratégique » transparait dans les discours américains. Le traitement des problématiques éthiques semble donc répondre, dans une certaine mesure, à la nécessité d'empêcher que l'intégration de l'IA dans l'appareil de défense ne soit entravée par ces questions, qui s'expriment déjà depuis de nombreuses années.

Au niveau international, la volonté de traiter les questions éthiques répond à plusieurs objectifs généraux. Comme l'*American AI Initiative*, stratégie nationale globale, l'indique clairement, ces efforts doivent permettre de façonner l'environnement international pour qu'il soit favorable aux intérêts américains. Il s'agit de mettre en place des cadres au niveau mondial qui aident notamment à la montée en puissance de la recherche nationale, accompagnent les performances des entreprises et groupes domestiques (notamment en laissant les marchés à l'exportation ouverts) et respectent les valeurs américaines. C'est en particulier dans la production des standards techniques internationaux que ces objectifs doivent être recherchés : ceux-ci doivent être favorables aux acteurs américains.

Le *narrative* fondé sur l'opposition entre pays démocratiques et nations qui utiliseraient l'IA de manière autoritaire et sans respect des droits de l'Homme peut ainsi être perçu, dans une certaine mesure, comme un médium permettant de faire du développement d'une IA éthique un intérêt et un objectif partagés par une communauté d'États globalement respectueux de certains corpus de droits et désireux de les protéger. Mais il sert également à justifier la mise en place de coopérations scientifiques et technologiques bilatérales ou multilatérales et d'une action commune pour l'édification des standards internationaux qui sont conçus – de manière assez explicite – comme des méthodes permettant aux entreprises américaines de dominer les marchés de l'IA¹.

Les mêmes logiques se retrouvent concernant l'IA de défense. Au niveau international, les États-Unis créent des partenariats et investissent les enceintes multilatérales en affirmant vouloir la mise en place de cadres pour s'assurer d'une IA militaire « responsable » – en opposition notamment à la Chine. Leur ambition est ainsi de stimuler, en particulier sur le plan de la R&D, les alliés et les partenaires. Les représentants américains présentent les lignes directrices éthiques produites par le pays et militent pour des procédures de sécurité partagées. Mais ces éléments permettraient également aux entreprises et aux armées américaines, déjà avancées dans ces domaines, d'imposer certaines normes à leurs partenaires, mais aussi concurrentes, et ainsi de disposer d'un avantage sur elles.

¹ Cette protection des intérêts nationaux peut même aller jusqu'à des tentatives de blocage d'initiatives internationales. Les États-Unis ont notamment fait évoluer certains projets dès lors qu'ils devaient aboutir à la création de normes « dures » (d'application obligatoire) et contenaient des dispositions pouvant avoir des conséquences défavorables pour les entreprises américaines.

Introduction

Les États-Unis ont été parmi les premiers pays à chercher à formaliser une position officielle et des politiques spécifiques à l'égard de l'Intelligence artificielle au cours des années 2010. Différents exercices de réflexion ont ainsi été menés, à un niveau global, concernant les apports potentiels de l'IA pour l'économie et la population américaines (et donc pas seulement pour la défense et la sécurité nationale). Les premiers documents généraux d'orientation traitant spécifiquement de l'IA ont été publiés à la fin du mandat de l'Administration Obama (2016).

L'architecture institutionnelle américaine a été très largement développée au cours des années 2018-2019. La politique fédérale en matière d'Intelligence artificielle progressivement définie repose sur une approche valorisant la pleine participation de tous les acteurs (entreprises, académiques, responsables politiques, organisations de consommateurs et autres représentants de la société civile). Il s'agit par ailleurs d'une politique publique cherchant à jouer sur de nombreux leviers. L'*Executive Order* ayant permis d'instaurer l'*American AI Initiative*, stratégie nationale globale, définit ainsi un large ensemble de domaines à améliorer (formation, cadre juridique, puissance informatique de calcul...) pour que la conception et la mise en œuvre des outils recourant à l'IA soient facilitées dans le pays. La politique de l'administration Trump semble toutefois caractérisée par le choix de privilégier, parmi tous les modes d'action, le soutien à la R&D, en particulier par l'octroi de financements².

Partout dans le monde, l'accélération du développement des technologies de l'Intelligence artificielle et leur diffusion progressive suscitent cependant de nombreuses interrogations. Les applications se multipliant dans des domaines de l'existence humaine très divers (sécurité, environnement, banque/finance, assurance, *marketing*, production manufacturière, santé...), ces technologies commencent en effet à modifier sensiblement les manières de travailler, d'apprendre et de vivre en société. Or, nombre d'acteurs (communautés de chercheurs, entreprises développant des applications, autorités politiques, associations et mouvements issus de la société civile...) voudraient notamment avoir la certitude que les transformations suscitées par l'Intelligence artificielle répondent bien à une approche humaniste, fondée sur les valeurs et les droits de l'Homme. D'autres – parfois les mêmes – s'intéressent aux changements que cette révolution technologique est en train d'avoir sur les rapports entre l'Homme et les machines.

Aux États-Unis, le traitement des questions éthiques que soulèvent les progrès de l'Intelligence artificielle, a été pleinement intégré aux politiques publiques lancées dans le domaine. Dès les premiers documents présidentiels d'orientation de 2016, des éléments d'analyse ont été proposés. Surtout, la réflexion sur ces thématiques a été définie comme une composante à part entière des actions publiques de soutien. Cette réflexion intègre la recherche de solutions concrètes (en particulier en matière technologique, juridique et de gouvernance) pour

² « United States of America. Defending the global AI leadership position », in Konrad Adenauer Stiftung, *Comparison of National Strategies to Promote Artificial Intelligence*, Part 1, 2019, p. 13.

Plus précisément, les autorités fédérales paraissent avoir fait de la promotion de la recherche fondamentale leur priorité. Cette conception des politiques en matière d'Intelligence artificielle est cohérente avec celle qui avait été initiée sous l'Administration précédente. Il s'agit d'employer les fonds publics de manière complémentaire à ceux engagés par les acteurs privés, qui sont essentiellement consacrés à la R&D orientée vers les applications.

que les risques identifiés avec le développement de l'IA ne puissent se matérialiser. Si l'Intelligence artificielle est donc définie dans tous les documents politiques d'orientation américains comme une révolution technologique que le pays ne doit pas rater – permettant d'améliorer la compétitivité des entreprises nationales, les capacités d'innovation de l'économie et renforçant la sécurité nationale –, le traitement des questions éthiques qu'elle soulève est considéré comme une condition indispensable pour que sa montée en puissance puisse être opérée.

La réflexion sur les questions éthiques que suscite le développement des emplois de l'Intelligence artificielle n'est bien évidemment pas propre aux États-Unis. Depuis le milieu des années 2010, institutions de recherche, organisations publiques et acteurs privés ont lancé, dans un certain nombre de pays, des exercices de réflexion pour constituer des principes et lignes directrices pour une IA éthique. Les débats portent notamment sur ce qui constitue l'éthique de l'IA, ainsi que sur les exigences éthiques, les standards techniques et les meilleures pratiques nécessaires à sa mise en œuvre. Ces efforts ont également été pris en charge par des organisations internationales, notamment des composantes des Nations Unies.

Pour tenter de répondre aux interrogations et aux peurs que suscite le développement de l'Intelligence artificielle, nombre de ces initiatives nationales et internationales reposent sur la constitution *ad hoc* de comités d'experts. Dans le cadre de leur saisine institutionnelle, ces organismes ont produit des rapports et des documents d'orientation concernant l'Intelligence artificielle. Des initiatives similaires ont été réalisées par le secteur privé, en particulier les grands groupes utilisant l'IA ou réalisant son développement commercial³. Des déclarations et des recommandations ont aussi été produites par des associations professionnelles, des alliances scientifiques et des organisations non-lucratives (comme *Amnesty International*)⁴. En 2019, les États-Unis et la Grande-Bretagne étaient les pays au sein desquels les documents produits – toutes origines confondues – étaient les plus nombreux⁵.

Des parties prenantes très variées sont donc impliquées dans les réflexions sur les questions éthiques que soulève l'actuelle montée en puissance de l'IA, et notamment dans la constitution de documents de cadrage. Chacune traite bien évidemment cette thématique de manière à composer avec ses propres priorités⁶. Les documents produits sont ainsi adressés à des publics parfois différents (certains étant même autodirigés, comme ceux rédigés par les fédérations scientifiques).

Dans le cadre particulier des applications de défense, ces questionnements éthiques prennent encore plus d'importance. La délégation aux machines, en particulier à des systèmes d'armes, de tâches de plus en plus nombreuses, réalisées de manière autonome, impose en effet de réfléchir à la possibilité que ces machines finissent par décider de la vie ou de la mort des

³ En 2018, *Google* et *SAP* – pour ne citer que ces exemples – ont rendu publics des lignes directrices et des principes.

⁴ Au milieu de l'année 2019, Jobin, Ienca et Vayena avaient ainsi identifié et comparé 84 documents nationaux et internationaux contenant des principes ou des lignes directrices éthiques pour l'Intelligence artificielle.

Jobin (A.), Ienca (M.) et Vayena (E.), « The global landscape of AI ethics guidelines », *Nature Machine Intelligence*, 1, 2019, pp. 389-399.

⁵ *Ibid.*

⁶ Greene (D.), Hoffmann (A. L.), Stark (L.), « Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning », *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

individus. Par ailleurs, l'émergence de dispositifs capables d'apprentissages autonomes et la complexification algorithmique croissante suscitent des problèmes particuliers puisqu'elles impliquent que les humains puissent ne pas comprendre sur quels fondements les décisions des systèmes sont prises.

Parmi d'autres, ces interrogations ont émergé avec la montée en puissance de l'IA militaire. Pour y faire face, les États-Unis ont assez largement adopté la même logique que pour la mise en place des politiques publiques plus globales en matière d'Intelligence artificielle. Si l'IA est considérée comme un ensemble de technologies que le DoD doit prioritairement intégrer – parce qu'elles ont la capacité potentielle de bonifier toutes les fonctions militaires (gestion des ressources humaines, entraînement, soutien médical, gestion des approvisionnements, entretien des équipements, renseignement, prise de décision...) et qu'une compétition au niveau mondial existe dans ce domaine scientifique et technologique – le traitement des questions éthiques est indispensable et ne peut être détaché des autres dimensions de la montée en puissance.

La présente étude a pour objectif de présenter les efforts des États-Unis pour traiter ces problématiques éthiques liées au développement, au déploiement et à l'utilisation par les armées des systèmes intégrant de l'Intelligence artificielle. Elle doit aussi replacer ces initiatives dans le cadre de l'écosystème national, mais également des discussions qui sont menées sur ces thèmes au niveau international. Pour ce faire, elle a été articulée en trois temps :

- ➔ La première partie est constituée d'un *mapping* des organismes qui traitent des dimensions éthiques de l'IA de défense aux États-Unis. Elle intègre également une présentation des documents institutionnels les plus importants produits par le pays dans ce domaine.

Cette étude du cadre institutionnel a été constituée à partir d'une logique en entonnoir. Sont d'abord analysés des éléments généraux, c'est-à-dire les cadres qui s'appliquent au niveau national pour toutes les politiques publiques de promotion de l'Intelligence artificielle. Dans un second temps, les dispositifs spécifiques à l'IA de défense sont présentés. Enfin, les organismes et documents traitant des questions éthiques liées aux applications militaires de l'Intelligence artificielle sont étudiés.

- ➔ La deuxième partie positionne les débats sur l'éthique de l'IA à un niveau international. Elle présente les principales initiatives globales lancées pour les organiser et obtenir des principes ou des lignes directrices (en précisant notamment les acteurs ayant initié ces efforts, l'organisation mise en place pour les prendre en charge et les productions réalisées).
- ➔ La troisième partie cherche à analyser le positionnement des États-Unis au sein de ces multiples cadres multilatéraux. Elle présente les enceintes que le pays a préférentiellement investies et les positions qui y sont soutenues.

Efforts de réflexion sur les questions éthiques au sein de l'écosystème américain

1. Cadre institutionnel général

1.1. Construction progressive du cadre institutionnel

1.1.1. 2016 et la pose des premières pierres de l'édifice

L'Administration Obama a lancé un effort de réflexion sur l'Intelligence artificielle. Celui-ci a abouti à des premières réalisations en 2016. Un *Subcommittee on Machine Learning and Artificial Intelligence* a notamment été formé au sein du *National Science and Technology Council* (NSTC). Entre autres fonctions, cette structure devait participer à la coordination des activités fédérales en matière d'IA.

Le rapport *Preparing for the Future of Artificial Intelligence* ainsi que le *National Artificial Intelligence Research and Development Strategic Plan* ont par ailleurs permis au NSTC de présenter, en octobre 2016, les tendances de la R&D de cette époque et de discuter des implications sociales et techniques du déploiement de l'IA à partir de quelques concepts clés (justice, responsabilité, sûreté et sécurité)⁷. Ces documents transverses ont aussi servi à proposer un certain nombre de recommandations en matière de recherche, de régulation et de politiques publiques de soutien par les administrations fédérales. Leur publication a permis de constituer un socle à partir duquel une stratégie nationale devait être produite.

Les efforts de réflexion n'ont pas émané que de l'Exécutif. En mai 2017, un *Caucus* bipartisan sur l'Intelligence artificielle a par exemple été constitué au sein du Congrès. Chargé d'informer les responsables politiques, notamment les parlementaires, sur les impacts technologiques, économiques et sociaux des avancées en IA, il devait constituer un outil permettant de mieux mesurer les bénéfices que les Américains pouvaient en attendre. En novembre 2017, il a notamment coorganisé un déjeuner-débat dédié aux questions éthiques et de protection de la vie privée.

Cette activité parlementaire aux multiples facettes s'est notamment traduite par le vote en décembre 2017 du *Future of AI Act*, à la fois à la Chambre des Représentants et au Sénat. Cette loi a surtout obligé le Secrétaire au Commerce à établir un *Federal Advisory Committee on the*

⁷ Un troisième rapport (*Artificial Automation, and the Economy*), rédigé par le *Council of Economic Advisors*, a traité spécifiquement de l'importance de l'AI et de l'automatisation pour la main d'œuvre américaine.

Development and Implementation of Artificial Intelligence, structure devant étudier des thématiques nombreuses et larges (compétitivité américaine, force de travail et délocalisation technologique, formation, questions éthiques, partage des données, coopération internationale, mise en responsabilité et droits, aspect culturel et normes sociales, impact sur les communautés rurales communautés et efficacité gouvernementale) et rendre compte aux parlementaires.

1.1.2. Renforcement des structures de coordination de haut niveau (2018-2019)

Les efforts engagés en 2015-2016 n'ont pas été prolongés lors des premiers mois de la présidence Trump. Aucun fonds spécifique n'a ainsi été créé pour l'Intelligence artificielle par les autorités fédérales. Le gouvernement n'a pas non plus produit de document politique central la concernant.

En mai 2018, la Maison Blanche a toutefois organisé un sommet national sur l'IA, assez largement ouvert puisqu'il a permis de rassembler une centaine de représentants de l'industrie, du monde académique et des administrations. Cet événement a constitué le début d'une nouvelle séquence. À cette occasion, Michael Kratsios, *Deputy Assistant to the President for Technology Policy*, a précisé l'approche développée par l'Administration Trump : maintenir l'avance américaine en matière d'IA ; soutenir les travailleurs américains ; alimenter la R&D publique ; faire disparaître les barrières à l'innovation ; rendre possible les applications spécifiques pour les industries.

Suite à ce sommet, un *Interagency Select Committee on Artificial Intelligence* a été créé. Rattaché au *National Science and Technology Council*, il a remplacé le *Subcommittee on Machine Learning and Artificial Intelligence* mis en place en 2016. Composé de responsables politico-administratifs fédéraux de haut niveau en charge des activités de R&D dans leurs organisations⁸, il est coprésidé par le *White House Office of Science and Technology Policy*, la *Defense Advanced Research Projects Agency* (DARPA) et la *National Science Foundation* (NSF). Son champ de compétences intègre donc les activités civiles comme militaires de recherche et développement.

Sa principale fonction est de fournir des expertises à la Maison Blanche sur les priorités générales en matière de R&D. Établi jusqu'à la fin de l'année 2020⁹, il doit également servir à faciliter la constitution de partenariats entre le gouvernement, le secteur privé et les communautés de chercheurs indépendants. Il s'agit d'une structure de coordination interministérielle et inter-agences de haut niveau, qui a aussi pour mission d'aligner les priorités des agences et Départements en matière de R&D et de vérifier la bonne exécution des planifications en matière d'investissements fédéraux.

⁸ Le directeur du *National Institute for Standards and Technology* (NIST), l'*Under Secretary for research and engineering* du DoD, l'*Under Secretary for science* du Département de l'Énergie, le directeur de la *National Science Foundation*, les directeurs de la DARPA et de l'IARPA...

⁹ Executive Office of the President, *Charter of the National Science and Technology Council Select Committee on Artificial Intelligence*, 2018.

Au cours de sa première réunion, en juin 2018, il a été annoncé la création de deux sous-comités :

- ➔ le *Machine Learning and AI (MLAI) Committee*, organe opérationnel de mise en œuvre des décisions de l'*Interagency Select Committee*. Il est composé de directeurs adjoints et sous-directeurs ;
- ➔ le *Networking and Information Technology Research and Development (NITRD) Committee*, un groupe de travail qui doit administrer la partie du programme NITRD dédié à l'IA¹⁰.

1.2. Publication de la Stratégie nationale en IA

En février 2019, un *Executive Order* a été rendu public, établissant une *American AI Initiative*¹¹. Institutionnalisant certaines des annonces qui avaient été réalisées lors du sommet national de mai 2018, le document est la stratégie attendue depuis la rédaction des documents généraux d'orientation politique de l'administration Obama. Elle avait été réclamée par de nombreux acteurs. La *National Defense Strategy (NDS)* de 2018 avait notamment regretté l'absence d'une référence commune pour l'ensemble des services gouvernementaux et des agences fédérales¹². L'*American Artificial Intelligence Initiative*, stratégie transverse coordonnée, a entre autres fonctions d'assurer l'amplification des efforts de R&D en IA et de favoriser les déploiements.

¹⁰ Le programme *Networking and Information Technology Research and Development*, lancé dans les années 1990, est le principal outil de financement fédéral sur les technologies de l'information avancées. Indépendant des outils de financement dont disposent les Départements et Agences, il permet à la Présidence de soutenir directement des projets de R&D et des actions permettant d'accélérer le transfert de la recherche vers la phase commerciale dans les domaines de l'informatique, des réseaux et des logiciels.

Le NITRD sert de réseau et permet notamment de mettre en contact permanent une douzaine d'agences fédérales en charge de la R&D sur les technologies de l'information et de la communication. La loi ayant créé le dispositif précise que les agences fédérales concernées posent collectivement les objectifs, pour chaque ensemble retenu (IA, cybersécurité, robotiques intelligentes et systèmes autonomes...), ainsi que les priorités en matière d'investissements et coordonnent les activités.

Le fait qu'une partie des fonds du NITRD soit gérée par une commission inter-agences dédiée à la R&D en Intelligence artificielle – et donc attribuée à sa montée en puissance – permet de constater que la Présidence fait véritablement de l'IA l'une de ses priorités et qu'elle intervient directement pour son soutien.

¹¹ Donald J. Trump, *Executive Order 13859. Maintaining American Leadership in Artificial Intelligence*, Washington D.C., The White House, February 11, 2019.

Un *Executive Order* est une directive émanant directement du Président des États-Unis. Il s'agit normalement d'un acte dirigé vers les agents gouvernementaux et les agences fédérales. Il vise à gouverner leur action (même s'il peut avoir des conséquences directes sur l'ensemble des citoyens), en leur imposant de prendre (ou de s'abstenir de prendre) certaines mesures, de modifier leurs pratiques, d'appliquer une politique publique déterminée par le Président ou d'exécuter la loi selon l'interprétation qu'il en retient. Pour être juridiquement valide, un *Executive Order* doit être rendu public (contrairement au *memorandum*).

¹² *National Defense Strategy of the United States of America. Sharpening the American Military's Competitive Edge*, January 2018.

Intitulé *Maintaining American Leadership in AI*, ce document rappelle l'importance de l'IA pour l'économie et la sécurité nationale. Cette stratégie nationale repose sur plusieurs ensembles d'actions :

► **Investir dans la R&D pour l'IA**

Les investissements fédéraux pour la R&D en Intelligence artificielle doivent augmenter. Ils devront être réalisés en collaboration avec les industries, la recherche académique, les partenaires internationaux et les alliés, ainsi qu'avec toutes les entités non-fédérales pouvant apporter un soutien, notamment des fonds.

Afin d'utiliser avec efficacité ces fonds, l'Administration a mis à jour son *AI R&D Strategic Plan*. De la sorte, les priorités en matière d'applications et de technologies ont été mieux présentées, permettant aux différentes parties prenantes de se positionner plus facilement par rapport aux intentions de l'État fédéral. Le Plan stratégique a d'ailleurs été accompagné d'un premier *reporting* des dépenses fédérales civiles pour la R&D en IA.

► **« Libérer » les ressources en IA**

La Stratégie nationale prévoit que les États-Unis augmentent l'accès aux données fédérales, ainsi qu'aux ressources en calcul de haute performance des acteurs fédéraux, afin de faciliter la R&D en IA¹³. L'*American AI Initiative* demande ainsi aux agences fédérales d'identifier les jeux de données d'intérêt pour la R&D qui ne sont pas encore libérés. Pour que ces nouvelles opportunités puissent être exploitées, elle leur demande d'étudier les méthodes pour pouvoir les transmettre aux acteurs qui pourraient légitimement les exploiter.

► **Faire disparaître les barrières à l'innovation en IA**

La Stratégie repose sur une réduction des barrières au développement, au test, au déploiement et à l'adoption sûre des technologies d'IA. Pour ce faire, elle demande que soient fournies des instructions sur la gouvernance de l'Intelligence artificielle, en accord avec les valeurs américaines, ainsi que des standards techniques appropriés.

Dans ce cadre, la Maison blanche a transmis aux administrations concernées (agences et ministères) un *Memorandum*, leur fournissant un premier jeu de lignes directrices pour concevoir les normes, d'application obligatoire ou volontaires, concernant l'IA¹⁴. De même, le *National Institute for Standards and Technology* (NIST) a été désigné pour rédiger la première Stratégie fédérale pour le développement des standards techniques en IA.

¹³ Tout en maintenant, voire en améliorant les protections concernant la sûreté, la sécurité, la vie privée et la confidentialité.

¹⁴ Memorandum for the Heads of Executive Departments and Agencies, *Guidance for Regulation of Artificial Intelligence Applications*.

► **Promouvoir un environnement international soutenant l'innovation en IA américaine**

Les acteurs nationaux, en particulier les autorités fédérales, doivent s'engager au niveau international pour promouvoir un environnement global qui soutienne la recherche et l'innovation en IA américaines et l'ouverture des marchés aux industries domestiques. Ils doivent également s'assurer que l'Intelligence artificielle soit développée dans des voies conformes aux valeurs nationales (en particulier en matière de protection de la vie privée, des droits et libertés civils). Plus globalement, le document appelle à configurer, au niveau international, l'Intelligence artificielle de manière à ce que son développement respecte les intérêts du pays.

► **Adopter une IA fiable pour les services et missions du gouvernement**

L'Initiative affirme que les administrations américaines, notamment fédérales, doivent adopter des outils dotés d'IA afin d'améliorer la fourniture des services au public, ainsi que leur efficacité (tout en améliorant la protection de la vie privée, les droits et libertés civils). Il s'agit aussi de stimuler la production d'IA en accentuant la demande émanant du secteur public.

Dans ce cadre, la *General Services Administration* a notamment créé un Centre d'excellence en IA, qui doit être capable d'identifier les meilleures pratiques développées au sein des agences fédérales pour intégrer à leurs organisations et services l'Intelligence artificielle.

► **Former une force de travail prête à exploiter l'Intelligence artificielle**

L'Initiative doit servir à renforcer les compétences des générations actuelles et futures de travailleurs grâce à l'utilisation de l'apprentissage, de programmes de formation et d'éducation aux sciences, technologies, ingénierie et mathématiques (*science, technology, engineering, and mathematics* – STEM). Un effort particulier doit être consenti sur les sciences informatiques.

Dans ce cadre, la Maison blanche a demandé à toutes les agences fédérales de rendre prioritaires les apprentissages, de même que les programmes professionnalisant, qui concernent l'Intelligence artificielle. Elle demande aussi de concentrer les financements fédéraux pour l'éducation vers les organismes ayant développé ou développant des cursus en IA.

Il est important de préciser que les autorités fédérales sont loin d'être les seules à avoir produit ou organisé des efforts de réflexion concernant l'IA, notamment les questions éthiques qu'elle génère. De nombreuses lois ont notamment été introduites dans les États fédérés. Leur préparation a bien évidemment suscité des débats. Les États fédérés ont été à l'origine d'autres types d'initiatives. En août 2018 par exemple, le Sénat de l'État de Californie a voté une résolution de soutien aux Principes pour l'IA d'Asilomar, un jeu de 32 lignes directrices portant sur le développement et l'utilisation sûrs et utiles de l'Intelligence artificielle inspiré des principes créés au niveau international en matière de recherche biologique. De même, le gouverneur de l'État du Vermont a décidé, en mai 2018, de créer une *task force* ayant pour mission de proposer des recommandations à propos des utilisations gouvernementales de l'IA et des régulations que les États fédérés pouvaient mettre en œuvre dans le domaine.

De même, de très nombreux textes portant sur les dimensions éthiques de l'IA – en particulier des codes et lignes directrices – émanant d'acteurs privés ont été progressivement constitués

dans le pays. Ils ont été produits par des fédérations professionnelles, des grands groupes industriels, des *think tanks*, des sociétés savantes... qui sont parfois reconnus par les autorités comme des centres d'expertise et participent donc aux débats publics, voire à la rédaction des documents officiels.

Figure n° 1 : PRINCIPALES INITIATIVES PRIVÉES EN MATIÈRE D'ÉTHIQUE DE L'IA

The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term	AI Now Institute	USA
Statement on Algorithmic Transparency and Accountability	Association for Computing Machinery (ACM)	USA
AI Principles	Future of Life Institute	USA
AI - Our approach	Microsoft	USA
Artificial Intelligence. The Public Policy Opportunity	Intel Corporation	USA
IBM's Principles for Trust and Transparency	IBM	USA
OpenAI Charter	OpenAI	USA
Our principles	Google	USA
Policy Recommendations on Augmented Intelligence in Health Care H-480 940	American Medical Association (AMA)	USA
Everyday Ethics for Artificial Intelligence. A practical guide for designers & developers	IBM	USA
Governing Artificial Intelligence. Upholding Human Rights & Dignity	Data & Society	USA
Intel's AI Privacy Policy White Paper. Protecting individuals' privacy and data in the artificial intelligence world	Intel Corporation	USA
Introducing Unity's Guiding Principles for Ethical AI – Unity Blog	Unity Technologies	USA
Digital Decisions	Center for Democracy & Technology	USA
Science, Law and Society (SLS) Initiative	The Future Society	USA
AI Now 2018 Report	AI Now Institute	USA
Responsible bots: 10 guidelines for developers of conversational AI	Microsoft	USA

AI Now 2017 Report	AI Now Institute	USA
--------------------	------------------	-----

Source : Jobin (A.), Ienca (M.) et Vayena (E.), « The global landscape of AI ethics guidelines », *Nature Machine Intelligence*, 1, 2019, pp. 389-399.

2. Cadre spécifique à la Défense

Dans le domaine de la défense, les États-Unis ont tout d'abord spécifiquement développé une démarche dédiée aux systèmes militaires autonomes, en particulier aux dispositifs armés. Pendant quelques années, l'approche américaine a même essentiellement reposé sur le concept d'autonomie. Dès 2012, la Directive 3000.09 (voir 1^{ère} Partie § 3.1.1.) – par la suite actualisée – portait sur le développement des plates-formes dotées d'une certaine autonomie. Dans ce cadre, l'IA ne faisait pas l'objet d'une approche spécifique¹⁵.

Plusieurs documents portant sur les utilisations de dispositifs autonomes au profit de la Défense ont ainsi été produits, notamment :

- ➔ U.S. Department of Defense/Office of the Secretary of Defense, *Unmanned Systems Integrated Roadmap: FY2017-2042*, June 2016 (1^{ère} édition 2011¹⁶) ;
- ➔ U.S. Department of Defense/Defense Science Board, *Summer Study on Autonomy*, June 2016 ;

¹⁵ Elle était essentiellement vue comme l'une des composantes techniques et technologiques du développement de l'autonomie des systèmes non habités et des plates-formes robotisées.

¹⁶ La première *Roadmap* a été publiée en 2011. Le document est actualisé environ tous les deux ans.

- ➔ U.S. Department of Defense/Joint Chiefs of Staff, *Joint Concept for Robotic and Autonomous Systems*, October 2016 ;
- ➔ U.S. Department of Defense, *Directive on Autonomy in Weapon Systems*, 3000.09, May 2017 (1^{ère} édition 2012).

Depuis quelques années, de très nombreux services et organisations du DoD portent cependant un intérêt grandissant à la conception et à l'utilisation d'équipements intégrant de l'Intelligence artificielle. Les programmes se sont ainsi multipliés¹⁷. Pourtant, jusqu'en 2018, il existait relativement peu d'outils de coordination. Le constat de ce manque s'est imposé, notamment lors de la rédaction de la *National Defense Strategy*. En mars 2018, dans une présentation devant une commission parlementaire, Mary Miller, *acting Assistant Secretary of Defense for Research and Engineering*, annonçait qu'une rationalisation des efforts internes au DoD en matière d'IA allait ainsi être menée.

2.1. Dispositions du National Defense Authorization Act (NDAA) 2019

Le NDAA FY 2019 a intégré différentes dispositions concernant l'Intelligence artificielle. Le texte pose comme objectif global au DoD de développer, améliorer et convertir les technologies de l'IA en vue d'emplois opérationnels. Pour y parvenir, il prévoit les fonctions de plusieurs autorités.

2.1.1. Nomination d'un haut responsable en charge de la coordination des activités de développement et démonstration des systèmes recourant à l'IA

La section 238 du NDAA donne pour mission à un haut responsable du ministère de coordonner les actions internes au DoD portant sur le développement de systèmes intégrant de l'IA ou du *Machine learning*.

Entre autres fonctions, il doit élaborer une feuille de route stratégique visant à l'identification et à la coordination des technologies de l'IA et des capacités critiques indispensables pour les soutenir (comme les capacités de calcul de haute performance). Par ailleurs, il a la charge de produire une évaluation continue des capacités en IA développées par le DoD, mais également à l'extérieur.

2.1.2. Création d'une Commission de sécurité nationale sur l'Intelligence artificielle (« AI Commission »)

Des dispositions, indépendantes de celles concernant le Coordinateur, ont permis la création d'une Commission de sécurité nationale sur l'Intelligence artificielle. Il s'agit d'un organisme de conseil indépendant mais rattaché au pouvoir exécutif¹⁸.

¹⁷ En avril 2018, Mickael Griffin, *under Secretary for research and engineering*, avait ainsi reconnu devant une commission parlementaire que 592 projets séparés impliquant de l'IA étaient à l'époque en cours de réalisation au sein du DoD.

¹⁸ La quinzaine de membres de la Commission est nommée par certains responsables politico-administratifs, notamment les ministres, et les présidents des commissions de la Chambre des Représentants et du Sénat. Pour une description un peu plus précise, voir 1^{ère} Partie, § 3.2.2..

La principale mission de l'*AI Commission* est de réaliser une revue, accompagnée de recommandations, pour le Congrès et les administrations. Elle a ainsi été créée pour proposer des méthodes et des chemins permettant de progresser dans le développement des outils d'IA, de *Machine learning* et des technologies associées répondant aux besoins de la sécurité nationale et des forces armées. Plus précisément, ses membres doivent apprécier la compétitivité des États-Unis dans ces domaines technologiques (en intégrant notamment les thématiques des liens public-privé et du niveau d'investissement en R&D). Ils doivent aussi analyser les tendances internationales (évolution des formes de coopération, montée en puissance de certains pays...). La Commission doit évaluer les risques associés aux emplois militaires de l'Intelligence artificielle et étudier les considérations éthiques et juridiques (notamment en droit international) afférentes. Son étude porte enfin sur les protocoles et standards, en particulier ceux devant être créés pour inciter à la libération des données indispensables à l'entraînement des systèmes.

Plus globalement, l'objectif attribué à l'*AI Commission* est de déterminer des méthodes favorisant le renforcement de l'attention que le pays accorde globalement à l'IA, au *Machine learning* et aux technologies qui leur sont associées pour des finalités liées à la Défense et à la sécurité nationale.

2.1.3. Renforcement des fonctions du Joint Artificial Intelligence Center (JAIC)

Les créations de l'*AI Commission* et du poste de Coordinateur IA viennent compléter celle du *Joint Artificial Intelligence Center* (JAIC). Celui-ci a été instauré pour coordonner, sur le plan technique, les actions des différents services du DoD gérant des programmes de développement d'outils recourant à l'IA.

Plus précisément, le JAIC a pour mission d'accélérer la livraison de capacités recourant à l'IA et de coordonner les différents organismes et services concernés par l'Intelligence artificielle au sein du *Department* (notamment synchroniser les programmes qu'ils lancent).

Contrairement à l'*AI Commission* et au Coordinateur, le JAIC existait avant la parution du projet de loi de finance. Une *Algorithmic Warfare Cross-Functional Team* avait été créée pour gérer le projet pilote MAVEN. Lancé en avril 2017, celui-ci avait pour fonction d'intégrer rapidement de l'IA au sein des systèmes existants du DoD, afin de démontrer le potentiel de la technologie. Le projet portait notamment sur l'exploitation des données issues des capteurs vidéo des drones.

Cette équipe est devenue le point focal, au sein du Département, pour l'intégration de l'Intelligence artificielle. En juin 2018, elle a ensuite constitué le noyau du JAIC lorsque celui-ci a été créé. Ses effectifs ont depuis été fortement augmentés (voir 1^{ère} Partie, § 3.2.3.).

2.2. Mission confiée au Defense Innovation Board (DIB)

Le Secrétaire à la Défense James Mattis a chargé, en juillet 2018, les experts du *Defense Innovation Board* d'étudier les dernières avancées en matière d'IA et de formaliser une réflexion sur les manières, pour les armées, d'employer ces technologies¹⁹. Le *Board* a notamment été chargé de susciter des discussions afin de créer des lignes directrices éthiques et un jeu de recommandations sur le développement et l'implémentation des technologies d'IA au sein du DoD.

2.3. Stratégie en IA du DoD

Adoptée à la fin de l'année 2018, la *Department of Defense Artificial Intelligence Strategy*, endossée par la Présidence, constitue l'un des principaux éléments de la rationalisation annoncée par Mary Miller, *acting Assistant Secretary of Defense for Research and Engineering*, en mars. En février 2019, le DoD a révélé l'existence du document, mais ne l'a pas publié (du fait de sa classification). Seul un résumé de 17 pages a été rendu public²⁰.

Le document permet de présenter l'approche stratégique développée par le DoD pour effectivement intégrer l'IA en son sein et ainsi maintenir les avantages militaires que le pays détient dans un certain nombre de domaines technologiques et opérationnels. Comme le résumé l'indique clairement, le Département considère que l'IA dispose de la capacité de bonifier toutes ses fonctions. La Stratégie ne porte donc plus exclusivement sur le développement et l'emploi des systèmes autonomes. Elle est plus large. Elle se situe dans la continuité du NDAA de juin 2018 dont elle reprend nombre de points.

La Stratégie est par ailleurs intimement liée à deux autres documents institutionnels. Elle s'inscrit tout d'abord dans le cadre de la *National Defense Strategy* de 2018. Dans sa présentation devant la presse, Dana Deasy, *Chief Information Officer* du Pentagone, a ainsi précisé qu'elle vient en soutien de cette dernière²¹. Les priorités posées par la NDS sont donc globalement les mêmes que celles de la *DoD Artificial Intelligence Strategy*.

Ensuite, la publication de l'*AI Strategy* a été concomitante de celle de l'*American Artificial Intelligence Initiative* (voir 1^{ère} Partie § 1.2.)²². Celle-ci présente un ensemble de demandes que l'Administration Trump fait aux agences publiques et aux différents départements ministériels. Dans une assez large mesure, l'*AI Strategy* du DoD est donc un document sectoriel applicatif de l'*Initiative*. Elle reprend et précise certains aspects de cette dernière pour la Défense.

La fonction première de la Stratégie est donc de créer un cadre pour le développement des applications impliquant de l'IA, quelles que soient leur temporalité et leur échelle. Il s'agit de disposer d'un ensemble de limites, objectifs et règles de gestion partagés, afin de bénéficier

¹⁹ Pour une rapide présentation du DIB, voir 1^{ère} Partie, § 3.2.1.

²⁰ Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy. Harnessing AI to Advance Our Security and Prosperity*.

²¹ Cronk (T. M.), « DOD Unveils Its Artificial Intelligence Strategy », U.S. Department of Defense, February 12, 2019.

²² Les publications des deux *Executive Orders* ont été réalisées à un jour d'intervalle.

d'une certaine unité. Ces lignes directrices sont conçues pour être communes à toutes les composantes du DoD (armées, organismes interarmées et agences). La Stratégie doit également permettre de rendre les processus d'acquisition de la Défense plus simples, adaptables (« agiles ») et rapides – les rédacteurs du document considérant que de nombreuses applications étant d'ores et déjà assez matures (ou en passe de l'être) pour être immédiatement (ou dans un avenir proche) intégrées dans les armées et les services du Département.

À plusieurs reprises, le résumé précise que l'introduction de l'IA a été pensée pour autonomiser et améliorer les capacités des personnels et non pour les remplacer. Il semble que la Stratégie ait aussi été conçue comme un outil devant permettre d'informer la communauté de la Défense et favoriser l'acceptation de l'IA en son sein.

La Stratégie suit une approche globale à 5 volets : améliorer la capacité à fournir des outils dotés d'Intelligence artificielle²³ ; maximiser l'impact de l'IA au sein du DoD²⁴ ; cultiver une force de travail spécifique ; stimuler les alliés et les partenaires ; diriger les efforts de réflexion en matière éthique et sur les questions de sécurité d'emploi.

Au niveau organisationnel, la Stratégie confirme le rôle central, attribué par le NDAA, au *Joint AI Center* au sein du DoD, notamment sa fonction de centre d'expertise pour l'ensemble des forces armées et des services du Département. Le Centre doit également fournir une vision commune à tous ces acteurs²⁵. Plus précisément, son action doit surtout permettre d'accélérer la livraison et l'adoption des systèmes recourant à l'Intelligence artificielle²⁶. Ses missions intègrent notamment la création d'un « socle commun » devant faciliter l'évaluation des impacts réels de l'IA dans les nombreux domaines de la Défense qui pourraient l'utiliser. Il s'agit notamment de concevoir des outils (cadres de référence, standards...). Son rôle est également de synchroniser les activités du DoD en matière d'Intelligence artificielle et de *Machine learning*, afin notamment de s'assurer qu'elles seront alignées avec les objectifs, priorités et temporalités définis dans la *National Defense Strategy*. Enfin, le Centre a pour fonction de lancer des actions permettant d'attirer et de générer des équipes en IA de niveau mondial au sein du Département.

L'action du JAIC vient donc compléter (et pas remplacer) celles de la DARPA, de la *Defense Innovation Unit* et des laboratoires des différentes armées.

Les alliés et même les partenaires des États-Unis sont cités à quelques reprises dans le document, tout comme les structures académiques. Dans les présentations officielles, il a été indi-

²³ Que ces outils aient été conçus en interne ou par des fournisseurs externes.

²⁴ Notamment permettre qu'un projet développé initialement pour un *Service*, le *Joint Staff*, un *Combatant Command*, une agence ou un bureau du Pentagone puisse facilement bénéficier à d'autres utilisateurs potentiels.

²⁵ Dans les *briefings* concernant la Stratégie, il a été confirmé que le JAIC peut intervenir à toutes les étapes du cycle de vie d'une application d'IA, même s'il doit concentrer ses efforts sur les stades tardifs de développement et sur l'adoption par les utilisateurs.

²⁶ Notamment en utilisant des solutions commerciales civiles existantes, adaptées aux emplois militaires (Simonite (T.), « The Pentagon Doubles Down on AI-And Wants Help From Big Tech », *Wired*, 12 February, 2019).

Lors d'une conférence ayant eu lieu à Baltimore en mai 2019 (l'AFCEA TechNet Cyber), Jill Crisman, membre du JAIC, indiquait ainsi qu'une division du travail avait été en quelque sorte mise en place : le JAIC s'occupe de l'« *AI now* », tandis que la DARPA se focalise sur l'« *AI next* ».

qué que les armées et les services du DoD ne pourraient parvenir à mettre en œuvre la Stratégie que grâce à des relations fortes avec leurs partenaires internes et externes²⁷. Il s'agit donc d'améliorer la coopération et les échanges entre les composantes du Département (JAIC, *Services, Joint Staff, Combatant Commands* et agences), mais également avec les acteurs académiques²⁸, les différentes agences publiques et les industriels. De même, la Stratégie implique que les liens avec les pays alliés et partenaires soient affermis.

L'expression « *non-traditional centers of innovation in the commercial sector* » est employée à quelques reprises dans le résumé public de la Stratégie. Cette idée est maintenant développée depuis de nombreuses années par les autorités. Celles-ci ont à plusieurs reprises répété qu'elles considèrent que, pour doter d'IA les armées et s'assurer de conserver une avance technologique dans ce domaine, les écosystèmes qui n'ont pas l'habitude de fournir des équipements et, plus largement, de travailler au profit de la Défense doivent être sollicités. La Stratégie doit ainsi reposer sur une multiplication des partenariats avec les industries, y compris celles qui ne sont pas en lien avec la Défense, afin que leurs dirigeants et cadres soient pleinement informés des défis que les armées veulent relever (et convaincus de les aider).

La Stratégie intègre aussi le fait de renforcer les liens entre les composantes du DoD et les différentes communautés de l'*open source*, et même d'accentuer la participation des personnels et des structures de la Défense au phénomène. Il s'agit donc d'amplifier les efforts qui avaient été formalisés avec la sortie d'un *mémoire* concernant l'*Open Source Software* (OSS)²⁹.

La Stratégie précise que des initiatives ont été et vont être lancées pour intégrer rapidement, mais de manière incrémentale, des systèmes recourant à l'IA pour les domaines principaux de missions des armées. Il s'agit notamment d'encourager les expérimentations devant être mises en œuvre rapidement (en quelques mois)³⁰. Ces projets devront permettre d'améliorer la connaissance situationnelle et la prise de décision, d'accroître la sécurité d'emploi des équipements, d'introduire de la maintenance prévisionnelle et de la gestion anticipée des approvisionnements, ainsi que de rationaliser les procédures commerciales du Pentagone et de ses agences.

De même, le document indique que les projets prioritaires sont ceux qui permettront le déploiement de systèmes déchargeant les personnels des tâches fastidieuses sur les plans physique et cognitif et introduisant de nouvelles méthodes de travail.

La Stratégie prévoit que le DoD investisse dans les recherches sur l'explicabilité de l'IA. Il doit également devenir pionnier dans le développement des outils de test, d'évaluation, de vérification et de certification. La finalité est en effet d'investir dans les travaux permettant de disposer de systèmes d'IA résilients, robustes, fiables et sécurisés. Le document reprend donc assez largement les objectifs de la Directive 3000.09 de 2012 (voir 1^{ère} Partie, § 3.1.1.).

²⁷ Cronk (T. M.), « DOD Unveils Its Artificial Intelligence Strategy », U.S. Department of Defense, February 12, 2019.

²⁸ Notamment en s'assurant que des financements suffisants permettront des recherches de long terme et la formation de spécialistes et de chercheurs de haut niveau.

²⁹ Department of Defense, Chief Information Office, *Clarifying Guidance Regarding Open Source Software*, Memorandum, October 16, 2009.

³⁰ Ces expérimentations doivent notamment servir à sensibiliser à l'IA les cadres, militaires et civils, du Pentagone, aider à la formation des personnels et favoriser leur acceptation des systèmes intelligents (les convaincre en quelque sorte).

Enfin, la Stratégie inclut une participation active du DoD, en particulier des militaires, aux réflexions éthiques sur l'IA. Il est ainsi prévu que des consultations soient organisées avec des membres du monde académique et des industries qui participent aux débats. Les composantes du Département doivent aussi pleinement collaborer aux discussions menées dans les enceintes internationales. Il s'agit de partager les lignes directrices éthiques et les procédures de sécurité qui vont progressivement être mises au point avec les autres nations. Il est toutefois intéressant de noter que le document, bien que faisant référence à la Directive 3000.09, comporte très peu de mentions des systèmes d'armes létaux autonomes (SALA).

2.4. Prolongements de 2019

L'année 2019 a été marquée par la publication d'autres documents. La Communauté du renseignement a notamment rendu publique, au début de l'année, une stratégie dédiée³¹. Elle prévoit une *Augmenting Intelligence using Machines (AIM) initiative* devant permettre, par des actions immédiates, de court, moyen et long termes, de modifier fondamentalement les méthodes et les logiques de production du renseignement.

Plus récemment, la *National Security Commission on Artificial Intelligence* a tenu sa première réunion plénière (mars 2019). Elle a poursuivi son travail et a publié, en juillet 2019, un rapport préliminaire adressé au Congrès. Un document intermédiaire a ensuite été transmis aux parlementaires en novembre.

2.5. Bilan

En dépit de la volonté de rationaliser les efforts nationaux, les décisions des années 2018-2019 ont amené la création ou la sollicitation d'un nombre relativement important d'organismes pour traiter des problématiques liées à l'Intelligence artificielle. Cette diversité concerne aussi bien les structures politico-stratégiques que techniques³². Elle se retrouve au sein même du DoD, où les organismes de conseil indépendants produisent des analyses, alors même que certains services sont également en charge des réflexions. Ce grand nombre de structures traitant des politiques de développement de l'IA a notamment pour conséquence de multiplier le nombre des documents institutionnels portant sur la thématique.

³¹ Office of the Director of National Intelligence, *The AIM Initiative: A Strategy for Augmenting Intelligence Using Machines*, Washington D. C., January 16, 2019.

³² Cette diversité correspond également à celle des organismes qui exécutent, gèrent ou supervisent des actions de recherche et développement en IA au sein du DoD. En plus du JAIC et de la DARPA, il convient en effet d'ajouter la *Defense Innovation Unit*, ainsi que les laboratoires de recherche des différents Services (comme l'*Air Force Research Laboratory* ou l'*Army Research Laboratory*) et *Combatant Commands*. Par ailleurs, certains organismes dédiés aux tests et évaluations sont également concernés par la montée en puissance des efforts en IA au sein du Département (comme l'*US Army Test and Evaluation Command*).

3. Efforts pour traiter les thématiques éthiques liées aux applications IA de défense

La montée en puissance de l'IA de défense aux États-Unis a été marquée par certaines tensions. Des débats sont nés il y a quelques années, qui ont notamment pris de l'ampleur lors de plusieurs séquences (salariés de *Google* menaçant de faire grève et exprimant leur refus que le groupe participe au projet Maven, sortie de la Stratégie IA du DoD...). Les critiques se sont assez largement concentrées autour de la thématique des SALA.

En mars 2019, le G^{al} Shanahan, commandant du JAIC, a ainsi reconnu que l'insuffisance de discussions publiques avait abouti à la diffusion de « graves idées fausses » sur les programmes menés au sein du Pentagone et, plus globalement, sur les intentions de la Défense dans ce domaine³³. Dans une certaine mesure, les différentes initiatives lancées pour nourrir les réflexions sur les problèmes éthiques que soulève l'utilisation de l'Intelligence artificielle par les armées ont, entre autres objectifs, pour fonction de rassurer.

3.1. Principaux documents dédiés spécifiquement aux dimensions éthiques

Susciter la réflexion sur les questions éthiques liées au développement et à l'emploi de l'IA de défense est un objectif qui a été affirmé dès le vote du NDAA en août 2018. Celui-ci allait même plus loin puisqu'il indiquait la volonté des autorités de disposer de documents cadres (lignes directrices et principes) en la matière. Ces éléments ont été repris dans la Stratégie du DoD. Le document fait même du traitement des questions éthiques l'un de ses cinq volets.

Cet objectif a connu un début de réalisation avec la publication d'un jeu de principes éthiques dédié à l'IA par le Pentagone.

3.1.1. Directive 3000.09

Comme indiqué précédemment, l'approche du Pentagone s'est essentiellement concentrée, pendant la première moitié des années 2010, sur la notion d'autonomie. C'est pourquoi la première version de la *Directive 3000.09*, qui traite de cette thématique, a été publiée en novembre 2012. Le document fournit en effet des lignes directrices concernant le développement des systèmes militaires armés autonomes. Elle pose un cadre qui doit être normalement respecté lorsque sont menées des actions de développement ou d'acquisition de ce type de dispositifs.

La *Directive* précise ainsi les responsabilités des différents services impliqués dans la conception, puis l'emploi, des systèmes d'armes autonomes ou semi-autonomes. Les lignes directrices qu'elle pose ont pour fonction principale de minimiser le risque de survenue de défaillances de ces systèmes (et de limiter les conséquences de ces pannes si elles surviennent), y compris lorsque les dispositifs sont employés lors d'engagements pour lesquels ils n'ont pas été initialement prévus.

³³ Williams (L. C.), « DOD AI chief: 'Grave misperceptions' about military programs », FCW, March 27, 2019.

La *Directive* fournit donc un cadre générique pour le développement de tous les dispositifs armés autonomes. Elle précise notamment que le Département de la Défense doit concevoir et utiliser des méthodes et outils, *software* comme *hardware*, de test et évaluation et de vérification et validation robustes. Ceux-ci doivent permettre une analyse des comportements émergents imprévus, résultant notamment de l'évolution des dispositifs autonomes et semi-autonomes dans un environnement opérationnel complexe. La *Directive* pose aussi que les services du DoD doivent être assurés que les systèmes fonctionneront comme anticipé lors des simulations et exercices, réalisés dans des environnements opérationnels réalistes (et notamment avec des ennemis adaptatifs). Enfin, il est affirmé que les dispositifs doivent être assez robustes pour minimiser les risques de pannes pouvant survenir lors d'engagements qui n'avaient pas été inclus initialement dans leurs domaines d'emploi. Dans ce cadre, la *Directive* insiste sur le fait que les interfaces homme-machine doivent être efficaces. Elles doivent permettre de rendre les systèmes autonomes facilement compréhensibles par des opérateurs entraînés, fournir des retours traçables sur le statut des systèmes et intégrer des procédures claires pour activer et désactiver les fonctions des systèmes (notamment les différents modes d'autonomie).

Plus globalement, la *Directive* stipule que, quel que soit le type de système, il doit être conçu de manière à permettre au commandement et aux opérateurs d'exercer les niveaux appropriés de jugement lors de l'utilisation de la force. L'application de ce principe implique donc d'exclure l'emploi de systèmes armés totalement autonomes. Surtout, tout système semi-autonome ou autonome doit pouvoir répondre à l'ensemble du processus d'évaluation des armements du DoD.

La *Directive* précise qu'en plus des autres procédures habituelles à respecter, les dispositifs autonomes et certains types de systèmes semi-autonomes doivent être approuvés, avant développement et avant déploiement, par l'*Under Secretary of Defense for Policy*, l'*Under Secretary of Defense for Acquisition, Technology, and Logistics* et le *Chairman of the Joint Chiefs of Staff*. Cet examen par des autorités de haut niveau ne concerne toutefois que les moyens aux fonctions les plus sensibles.

Le cadre que ce document permet de poser n'est pas exempt de limites. Certains observateurs ont dénoncé le manque de précision de la *Directive*. Les « *niveaux appropriés de jugement humain concernant l'emploi de la force* » ne sont en effet pas spécifiés³⁴. Le DoD considère en réalité que le terme « approprié » implique une flexibilité et que les niveaux de jugement des opérateurs humains ne vont pas être toujours les mêmes (variant selon le type de système d'arme, la situation opérationnelle, le type d'IA employée...). Cette position a d'ailleurs été clairement présentée dans un document datant d'août 2018, qui complète en quelque sorte la *Directive*³⁵ :

« *'appropriate' is a flexible term that reflects the fact that there is not a fixed, one-size-fits-all level of human judgment that should be applied to every context. What is*

³⁴ Voir Saxon (D.), « A Human Touch: Autonomous Weapons, DoD Directive 3000.09 and the Interpretation of 'Appropriate Levels of Human Judgment over the Use of Force' », in Bhuta (N.), Beck (S.), Geiß (C.), eds., *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge, Cambridge University Press, 2016.

³⁵ U.S. Government, « Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, » August 28, 2018.

‘appropriate’ can differ across weapon systems, domains of warfare, types of warfare, operational contexts, and even across different functions in a weapon system. »

Plus globalement, la Directive pose donc un certain nombre d'exigences, mais ne précise pas concrètement (notamment techniquement) comment y répondre. Certains observateurs, en particulier des experts de *think tanks*, se sont ainsi interrogés, à la fin de l'année 2019, sur la nécessité de mettre à jour ou de compléter la Directive³⁶. Quelques-uns d'entre eux ont notamment affirmé que, seule, elle ne pouvait suffire comme socle de la politique du Pentagone concernant les limites éthiques à l'emploi des systèmes d'armes autonomes. Ils considèrent notamment que la Directive n'est pas assez large pour intégrer les développements technologiques en cours.

En septembre 2020, lors d'un *briefing* public sur les travaux menés par le JAIC, des représentants du Centre ont réaffirmé que les activités de la structure étaient menées en respectant les cadres, notamment juridiques, s'appliquant au DoD et que – conformément à la Directive 3000.09 – aucun système armé véritablement autonome n'est déployé par les armées américaines³⁷.

3.1.2. Principes éthiques du DoD

En février 2020, le DoD a annoncé la publication d'un jeu de principes éthiques spécifique à l'Intelligence artificielle. Il s'agit d'un complément à la Stratégie du Département (voir 1^{ère} Partie, § 2.3.). Il doit permettre de réaliser les objectifs de la Stratégie tout en assurant un emploi légal des systèmes dotés d'Intelligence artificielle. Surtout, ces principes sont censés permettre aux États-Unis de prouver qu'ils sont à l'avant-garde des réflexions sur ces questions.

Le document a été conçu en suivant les premières recommandations du *Defense Innovation Board* (octobre 2019) et les propositions contenues dans le rapport intermédiaire de l'*AI Commission* (octobre 2019). Il a été produit à partir d'un travail de 15 mois, fondé notamment sur des consultations avec des experts de l'IA issus des industries, des milieux académiques et des administrations.

A. Contenu du document

L'objectif avec la publication des Principes est de fournir un « vocabulaire partagé » à tous ceux qui travaillent au développement des technologies – des personnels en charge de la rédaction des marchés publics aux concepteurs de logiciels – afin que les dimensions éthiques soient prises en compte dans leur travail.

Les principes n'ont pas vocation à s'appliquer uniquement aux systèmes armés. Ils ont en effet été conçus pour concerner aussi bien les applications de combat que les systèmes non directement affectés aux fonctions de combat ou à leur soutien. Ils ont été rédigés de manière à

³⁶ Atherton (K. D.), « When should the Pentagon update its rules on autonomous weapons? », C4ISRNet, December 12, 2019.

³⁷ Vincent (B.), « Joint Artificial Intelligence Center officials stay connected with stakeholders through an executive steering group and other working groups », Nextgov, September 10, 2020.

respecter plusieurs ensembles de droit : la Constitution, le Titre 10 de l'*United States Code*³⁸ (qui fournit le cadre légal pour les missions, rôles et organisations des forces armées), ainsi que le droit international applicable aux conflits armés et les traités internationaux.

Il constitue donc un cadre global pour tous les types d'applications. Lignes directrices de haut niveau, ce cadre doit être respecté par toutes les agences du DoD et par les *Services* dès lors que ces acteurs mettent en œuvre des dispositifs recourant à de l'Intelligence artificielle.

Les principes sont au nombre de 5.

▶ **Un usage responsable**

Le personnel du DoD doit exercer des niveaux appropriés de jugement et d'entretien de manière à toujours demeurer responsables lors du développement, du déploiement et de l'utilisation des capacités en IA.

▶ **Un usage équitable**

Le DoD doit volontairement franchir des étapes prédéfinies pour minimiser le risque de développement de biais inattendus lors de l'exploitation des capacités en IA.

▶ **Un usage tracé**

Les capacités en IA du DoD doivent être développées et déployées seulement si les personnels impliqués possèdent une compréhension appropriée de la technologie, des processus employés pour sa conception et des méthodes opérationnelles applicables aux capacités qu'elle soutient. Cet objectif implique que les méthodologies, l'origine des données, les procédures de *design* et la documentation soient transparentes et auditables.

▶ **Un usage fiable**

Les capacités en IA du Département de la Défense doivent faire l'objet d'utilisations explicites et parfaitement définies. La sécurité, la sûreté et l'efficacité de ces capacités doivent avoir été sujettes à des tests et des assurances, qui concernent les usages définis pour l'ensemble des cycles de vie des systèmes.

▶ **Un usage gouvernable**

Le Département doit concevoir et produire des capacités en IA qui remplissent les fonctions que l'on attend d'elles tout en détenant la possibilité de détecter et éviter les conséquences inattendues et de désengager les systèmes déployés dès lors qu'ils ont des comportements non prévus.

L'adoption des Principes reflète le choix du Pentagone de ne pas limiter les problèmes éthiques liés à l'adoption de l'IA aux seuls SALA. Sur les 5 Principes, 4 concernent les risques relatifs aux biais, le besoin d'une supervision humaine à tous les stades de conception de la

³⁸ Le texte est une codification du droit américain fédéral à caractère général et permanent. Cette codification est effectuée par l'*Office of the Law Revision Counsel* de la Chambre des représentants.

militarisation de l'Intelligence artificielle et la capacité à identifier et éliminer les défauts. Seulement un Principe (la gouvernabilité) traite véritablement du risque d'actions inattendues et potentiellement létales par des systèmes autonomes dotés d'IA.

B. Limites des Principes

Les Principes recourent à certains termes – fondamentaux – qui ne sont pas définis avec précision. Le principe de responsabilité repose sur un niveau « approprié » de contrôle. Comme pour la Directive 3000.09, certains observateurs ont fait valoir qu'il faudrait sans doute que le qualificatif « approprié » bénéficie de clarifications. De même, le principe d'équité impose une réflexion sur les « biais inattendus » et leur traitement, sans que ceux-ci ne soient pourtant définis ou catégorisés.

Par ailleurs, les Principes constituent des lignes directrices générales. Leur mise en œuvre implique qu'ils soient précisés. Ils indiquent ce que les services et agences du DoD doivent rechercher et respecter, mais pas comment ils peuvent y parvenir.

Pour qu'un usage responsable de l'IA puisse être assuré, il semble par exemple nécessaire de distinguer plusieurs éléments³⁹. Au cours de la conception du système, la responsabilité peut signifier de définir relativement précisément les limites d'emploi et de s'assurer que ces limites soient toujours respectées. Ceux qui testent ensuite le dispositif doivent déterminer les frontières « basses » et « hautes » du domaine d'emploi, de manière à ce que la sécurité puisse être assurée.

Dès lors que le système est sur le terrain, le respect de la responsabilité implique de rendre compte – ce qui ne peut être effectué que si une chaîne claire de commandement existe. Il s'agit notamment de toujours savoir qui sont les opérateurs et les décisionnaires qui choisissent de lancer les actions, à partir notamment des informations transmises par le système recourant à l'IA.

Le caractère équitable n'est en réalité possible que si l'idée que les fonctionnements mentaux et sociaux humains sont sous-tendus par de nombreux biais est intégrée. Les concepteurs des systèmes reposant sur du *Machine learning* doivent notamment toujours prendre en considération que ceux-ci apprennent à partir des jeux de données qui leur sont fournis par les Hommes. Il faut donc chercher et éliminer les biais que ces jeux de données contiennent. Pour les découvrir, il faut que soient mises en place des méthodologies spécifiques, qui ne sont pas décrites – même brièvement – dans le document.

La traçabilité repose sur le fait qu'une documentation transparente sur les cycles de vie des modèles d'IA soit accessible si besoin⁴⁰. L'identification des modèles et des différentes versions de chaque modèle constitue une tâche qui pourrait aider les agences du DoD à s'assurer que les résultats obtenus lors du développement d'un système doté d'IA puissent être répliqués. Elle permettrait également de rendre l'altération du système plus difficile.

³⁹ McPherson (S.), « Ensuring the Pentagon follows ethics for artificial intelligence », C4ISRNet, May 3, 2020.

⁴⁰ Les cartes que Google a rendues publiques en 2019 représentent un exemple de documentation. Elles aident à définir le contexte dans lequel chaque modèle d'Intelligence artificielle est censé être utilisé et offre des détails sur les procédures d'évaluation des performances – ce qui accroît la transparence.

Le suivi pourrait permettre de développer la fiabilité. Le DoD définit la fiabilité comme des « usages explicites, parfaitement définis », accompagnés « de la sûreté, de la sécurité et de l'efficacité ». L'un des aspects principaux de cette définition de la fiabilité est la capacité à se prémunir contre toute dérive du modèle d'IA employé par le système. De nombreux modèles fonctionnent parfaitement avec des données d'entraînement. Dès qu'ils emploient des données réelles, notamment issues de différentes sources, leur précision chute. Les acteurs en charge du développement et de l'emploi des systèmes doivent donc suivre leurs statistiques, ainsi que celles des senseurs utilisés pour collecter les données – ce qui implique la mise en place de procédures et d'outils dédiés.

Le caractère gouvernable d'un dispositif militaire recourant à de l'Intelligence artificielle impose notamment que l'utilisateur ou le superviseur soit capable de déterminer lorsque le système ne fonctionne pas comme prévu et puisse le désengager. Le DoD n'indique pas dans les Principes que le caractère gouvernable peut être assuré dès la conception des dispositifs – alors qu'il semble que ce soit l'une des voies les plus prometteuses à travailler (*governability by design*).

Enfin, la pleine application des Principes requiert que tous les personnels impliqués dans la conception, le déploiement et l'emploi des systèmes recourant à de l'IA les connaissent et soient formés à les utiliser concrètement. Il en est de même des entreprises fournissant les dispositifs au DoD.

3.2. Principales structures impliquées dans le traitement des dimensions éthiques de l'IA de défense

3.2.1. Le Defense Innovation Board

Le *Defense Innovation Board* a été instauré en 2016. Rassemblant 10 membres issus essentiellement des communautés de chercheurs/enseignants-chercheurs et de secteurs industriels de pointe, il s'agit d'un organisme devant fournir au Secrétaire à la Défense et aux principaux responsables du DoD des analyses et des recommandations indépendantes sur les technologies émergentes et les approches innovantes qui pourraient être adoptées par le Département, afin que les États-Unis conservent leur supériorité technologique et militaire⁴¹. Les travaux du *Board* servent également à nourrir les processus législatifs. Ses recommandations doivent normalement être relativement concrètes et pouvoir être mises en œuvre rapidement.

L'expertise que le *Board* fournit peut porter aussi bien sur les dimensions humaines (formation, culture...), les technologies et les capacités militaires qu'elles servent à entretenir, que sur les pratiques et les changements organisationnels nécessaires pour exploiter les nouveautés. Le DIB est ainsi constitué de 3 sous-comités : *Science & Technology* ; *Workforce, Behavior and Culture* ; *Space Advisory Committee*.

⁴¹ Actuellement, le DIB travaille plus particulièrement sur l'IA, les données et leur exploitation, l'algorithmie et les logiciels, la transition digitale et le capital humain nécessaire pour développer et exploiter ces domaines technologiques, ainsi que sur les questions spatiales.

Dès janvier 2017 et la publication de ses premières recommandations pour renforcer les capacités d'innovation au sein du DoD, le DIB a produit une proposition spécifique concernant l'Intelligence artificielle⁴². Comme indiqué précédemment, le Secrétaire à la Défense James Mattis a par ailleurs mandaté, à l'été 2018, les experts du *Defense Innovation Board* pour qu'ils formalisent une réflexion sur les manières, pour les armées, d'employer l'IA et les technologies qui lui sont liées. Dans le cadre de cette réflexion (à laquelle le nom d'« *Artificial Intelligence Principles Project* » a été donné), le *Board* a notamment été chargé de créer des lignes directrices éthiques, ainsi qu'un jeu de recommandations sur le développement et l'implémentation des technologies d'IA au sein du DoD.

En octobre 2018, les membres du *Board* ont fait savoir qu'ils comptaient finir leurs travaux dans les 9 mois. Début 2019, le DIB a sollicité un mix de chercheurs (notamment académiques), de spécialistes de l'éthique, de juristes, de dirigeants d'entreprises, d'experts en politique publique, de représentants d'associations et de responsables de structures de capital-risque pour qu'ils fournissent des réflexions pour nourrir ses travaux. En février 2019, le *Science & Technology Committee* du DIB a également organisé une session publique, à l'Université de Stanford, au cours de laquelle il a été possible à des citoyens de poser à l'avance des questions aux panélistes, puis de commenter leurs présentations⁴³. Les auditions et interactions produites au cours de la session publique ont servi d'*inputs* pour l'étude commandée par le Secrétaire à la Défense.

En octobre 2019, le *Defense Innovation Board* a finalement présenté plusieurs documents lors d'une réunion publique. Deux d'entre eux (un document maître et un autre rassemblant des annexes) proposaient un jeu de principes éthiques concernant la conception et l'emploi de l'Intelligence artificielle pour les applications défense⁴⁴.

3.2.2. L'AI Commission

Comme indiqué précédemment, la *National Security Commission on Artificial Intelligence* a été établie par le NDAA FY19. Organisme indépendant, elle est simplement soutenue par le DoD pour certains problèmes administratifs : le Département lui fournit ainsi une équipe de soutien, des locaux et contracte pour elle lorsque cela s'avère nécessaire.

Composée de 15 membres⁴⁵, la Commission doit, entre autres missions, évaluer les risques associés aux emplois militaires de l'Intelligence artificielle et étudier les considérations éthiques et juridiques (notamment en droit international) afférentes (objectifs 6 et 7 de la charte constitutive). Son regard doit aussi porter sur les protocoles et standards devant être créés pour inciter à la libération des données indispensables à l'entraînement des systèmes.

⁴² La recommandation 5 (« *Catalyze Innovations in Artificial Intelligence and Machine Learning* »), qui comportait plusieurs propositions.

La décision de créer le JAIC semble ainsi avoir été largement inspirée par l'une de ces propositions qui recommandait d'établir un centre au sein du DoD pour l'étude de l'IA et du *Machine learning*.

⁴³ Une retranscription de cette conférence publique est accessible sur le site du DIB.

⁴⁴ Defense Innovation Board, *AI Principles: Recommendations on Ethical Use of Artificial Intelligence by the Department of Defense*, 2019.

⁴⁵ Dont des représentants de quelques-unes des principales entreprises du domaine (*Oracle, Microsoft, Amazon Web Services* et *Google*).

Dans son rapport intermédiaire (octobre 2019), la Commission proposait quelques premiers développements sur les questions éthiques. Dans la version de février 2020, les réflexions ont été plus étoffées. Ces éléments avaient été préalablement transmis au JAIC pour qu'ils puissent les intégrer dans les Principes éthiques en cours de rédaction. Finalement, un document dédié a été proposé par la Commission lors de la publication de ses travaux finaux en juillet 2020⁴⁶.

3.2.3. Le JAIC

Comme indiqué précédemment, le noyau du JAIC était composé de l'*Algorithmic Warfare Cross-Functional Team*, créée pour gérer le projet pilote MAVEN (lancé en avril 2017). Pendant les premiers mois de son existence, le Centre a essentiellement fonctionné avec les membres de cette équipe. À la fin de l'année 2018, six mois après sa création, le JAIC disposait ainsi d'effectifs encore réduits, composés majoritairement d'officiers détachés pour 6 mois de leurs unités ou services d'appartenance. Depuis, le nombre de ses personnels a crû en permanence. Pour l'année budgétaire 2020, il est prévu que le Centre dispose de 75 personnels.

Le JAIC est l'un des services du *Department of Defense Chief Information Officer*⁴⁷. Depuis sa naissance, l'organisme a été progressivement structuré. Trois nouveaux postes ont notamment été créés (*chief science officer, chief technology officer, chief of acquisitions*), permettant de renforcer l'encadrement. Ces postes ont par ailleurs été confiés à des personnes disposant d'une expérience (industrielle, académique et administrative) significative.

Le JAIC a joué un rôle majeur dans la génération de la Stratégie du DoD en matière d'IA. Il a également été fortement impliqué dans la conception des Principes éthiques, en particulier dans la rédaction du document. Pour ce faire il a beaucoup échangé avec le DIB, les exercices étant assez largement croisés.

En septembre 2020, des représentants du *Center* ont rappelé que le développement, en lien avec les armées, de lignes directrices éthiques constitue toujours l'une des priorités de la structure⁴⁸. Les travaux en son sein continuent sur cette thématique.

Afin de coordonner (surtout synchroniser) les actions des différentes composantes du DoD travaillant à la conception et au déploiement de l'IA de défense, différents comités ont été constitués – certains comportant des sous-comités⁴⁹. Ces groupes de travail, qui se réunissent régulièrement, constituent des cadres pour que les spécialistes du Centre et les responsables du Département non affectés au JAIC puissent se rencontrer et travailler ensemble⁵⁰. Ils trai-

⁴⁶ National Security Commission on Artificial Intelligence, *Key Considerations for Responsible Development & Fielding of Artificial Intelligence*, July 22, 2020.

⁴⁷ Alors que l'*Algorithmic Warfare Cross-Functional Team* était placée sous l'autorité de l'*Under Secretary of Defense for Intelligence*.

⁴⁸ *Ibid.*

⁴⁹ Vincent (B.), « Joint Artificial Intelligence Center officials stay connected with stakeholders through an executive steering group and other working groups », Nextgov, September 10, 2020.

⁵⁰ Ces groupes de travail interarmées et interservices sont intégrés au travers d'une procédure de supervision par un *executive steering group* (ESG) réunissant des responsables de haut niveau (« *three star* ») des *Services, com-*

tent de thématiques spécifiques. L'un d'entre eux est responsable des travaux sur les problématiques éthiques ; un autre de la mise au point des cadres pour les tests et évaluations⁵¹. Ces deux comités travaillent notamment à la rédaction d'instructions pour la passation des marchés publics, ainsi qu'à la conception de garanties technologiques, de contrôles organisationnels, de stratégies de limitation des risques et de mesures de formation.

Le JAIC a été désigné pour être le point focal, au sein du DoD, pour la coordination de la mise en œuvre des Principes éthiques en matière d'IA. Dans ce cadre, il a lancé 4 premières actions.

- ▶ **Les deux premières sont d'une part une réflexion pour faire des principes des standards utilisables dans les appels à propositions du DoD, d'autre part la mise de la thématique de l'application des Principes à l'agenda du groupe de travail interservices et interarmées dédié aux questions éthiques que le Centre soutient.**

L'objectif est de publier, grâce à un travail interservices et interarmées, un document de référence plus large que les seuls principes. Dans un premier temps, une équipe au sein du JAIC a travaillé seule. Un groupe de travail « *Responsible AI* » a par la suite été créé au sein du sous-comité interservices dédié aux questions éthiques.

Ce document devra servir de guide de référence pour l'ensemble du Département. Il permettra d'expliquer les 5 Principes et de décrire des méthodes pour les mettre en œuvre lors de la conception et l'emploi des systèmes utilisant de l'IA.

- ▶ **La troisième a trait au lancement du programme *Responsible AI Champions*.**

L'équipe interne du JAIC a travaillé à la définition de programmes pilotes de formation en éthique. Un premier projet a commencé à être mis en œuvre. Cette action consiste à former des individus aux problèmes éthiques grâce à un cycle de réflexion d'environ deux mois⁵². La première vague de « champions » a été constituée au début du mois de juillet 2020. Il s'agissait de 15 membres du JAIC, issus de toutes ses divisions fonctionnelles. L'objectif est à terme de constituer une communauté de personnels compétents capables de « défendre » les Principes éthiques et d'aider à définir les manières de les mettre en œuvre concrètement (en particulier dans les procédures d'ingénierie). Chacun des « champions » formés va pouvoir diffuser cette sensibilité pour les questions éthiques au sein de l'équipe à laquelle il appartient et même être engagé dans des méthodes d'apprentissage de pair-à-pair. Par ce dispositif, il s'agit

batant commands et des différents organismes du DoD en charge de la R&D et de l'ingénierie. Toutes les composantes du Département concernées par l'IA sont ainsi représentées au sein de ce groupe de suivi, qui sert de forum.

⁵¹ Vincent (B.), « Joint Artificial Intelligence Center officials stay connected with stakeholders through an executive steering group and other working groups », Nextgov, September 10, 2020.

Afin de répondre aux exigences de la Directive 3000.09, le JAIC est notamment en contact – en particulier au travers de ce comité dédié – avec tous les organismes responsables des tests et évaluations (dont l'*Office of the Secretary of Defense for policy*). Il dispose des compétences nécessaires pour jouer la fonction de *red team* lors des procédures qu'ils emploient.

⁵² Neuf semaines sont consacrées à un apprentissage expérientiel servant à nourrir des discussions autour des voies utilisables par le DoD pour appliquer concrètement les Principes, mais également celles qui peuvent être développées par chaque participant dans le cadre de ses fonctions.

Le cycle est animé par une équipe pluridisciplinaire d'experts thématiques qui recourent à des supports d'enseignement, des débats et des études de cas.

donc de passer du concept à des pratiques concrètes, utilisables lors des procédures qui permettent l'acquisition, le développement, le déploiement et même le soutien des systèmes utilisant de l'Intelligence artificielle.

Depuis l'été 2020, le JAIC a associé à la conception du programme *Responsible AI Champions* d'autres composantes du DoD et des *Services*. L'objectif semble en effet être que les auditeurs des futures promotions soient issus de tout le Pentagone.

Le programme des « Champions » et la rédaction du guide devront donc permettre de disposer d'un vocabulaire partagé au sein du Pentagone et de rendre « naturelle » la pratique des garde-fous éthiques, de même que la réflexion sur eux⁵³.

► **La quatrième réside dans la participation aux travaux sur la création d'un *Data Governance Council* au sein des administrations.**

Dans les textes institutionnels produits par la Défense américaine, un lien est très clairement fait entre nombre de questions éthiques et les problèmes de test et d'évaluation des systèmes dotés d'IA (voir ci-après). Le JAIC a donc développé une stratégie concernant les tests, qui cherche à jouer sur plusieurs registres⁵⁴. La création de listes de paramètres prédéfinis pour évaluer les technologies en IA est l'un des travaux engagés. De même, le suivi et l'évaluation continus des modules d'IA intégrés dans des systèmes et plateformes existants est également une action majeure dans le cadre de cette stratégie. L'organisation de tests opérationnels directs est aussi recherchée. Le JAIC attend par ailleurs que des outils de test produits par le secteur privé soient disponibles, afin de les évaluer et de définir comment ils pourraient être adaptés pour les systèmes militaires.

La démarche actuellement mise en œuvre par le JAIC pour mettre en place une éthique dédiée à l'IA de défense consiste donc à jouer à la fois sur la composante humaine (notamment par la sensibilisation/formation) et les outils digitaux⁵⁵.

4. Thématiques sous-tendant le traitement des questions éthiques

Lors de la présentation des Principes éthiques du DoD en février 2020, le Secrétaire à la Défense, Mark Esper, a rappelé que l'intégration de l'Intelligence artificielle et des technologies qui lui sont associées était l'une des priorités du Département. Dans une certaine mesure, positionner le Département à l'avant-garde du traitement des questions éthiques semble donc être une condition facilitant cet objectif (notamment au travers d'un renforcement de l'acceptation et par l'atténuation des peurs et critiques).

Dès le NDAA FY 2019, l'objectif d'intégrer rapidement l'IA au sein du DoD, en particulier des armées, a en effet été affirmé. Comme beaucoup d'autres pays, les États-Unis partent du cons-

⁵³ Barnett (J.), « How the DoD is developing its AI ethics guidance », fedscoop, September 2, 2020.

⁵⁴ *Ibid.*

⁵⁵ *Ibid.*

tat que l'IA est un ensemble de technologies génériques qui auront des conséquences majeures dans tous les secteurs de la vie humaine (des « *General Purpose Technologies* »), renforçant notamment les capacités de la défense nationale et de l'économie. Il s'agit donc d'accélérer un mouvement déjà engagé : de nombreuses technologies étant désormais assez matures, il faut le plus rapidement possible achever le développement d'applications, les tester et les intégrer aux systèmes et organisations existants. L'objectif n'est donc plus seulement de renforcer la R&D, mais bien de disposer de systèmes opérationnels, y compris pour des fonctions centrales pour les missions des armées sur le terrain.

Dans les discours américains, en particulier les documents institutionnels, l'attention pour l'IA est également justifiée par l'idée qu'une compétition dans ce domaine se développe à l'échelon mondial. Les efforts russes et surtout chinois sont ainsi très souvent cités. Dans ce cadre, il s'agit donc non seulement d'assurer le renforcement de l'efficacité des armées, mais surtout de maintenir l'avance dont les États-Unis disposent dans ces domaines technologiques – multiplicateurs de puissance militaire – face aux avancées d'adversaires. La Stratégie du DoD s'inscrivait ainsi dans le cadre global de la *Third Offset Strategy* (TOS) de l'Administration Obama, dont la logique perdure⁵⁶. Elle est caractérisée par l'expression d'un certain sentiment d'urgence face à un risque de « déclassé stratégique ».

La reconnaissance par le Pentagone de l'importance des dimensions éthiques appliquées aux usages militaires de l'IA répond notamment à trois jeux principaux de préoccupations qui ont émergé ces dernières années⁵⁷ :

- ➔ Le premier jeu regroupe les peurs concernant une baisse de la surveillance que l'Homme est capable d'opérer sur les systèmes (et ainsi que les dispositifs dotés d'autonomie puissent développer des comportements « voyous », notamment en violant le droit international humanitaire) ;
- ➔ Le deuxième rassemble les craintes concernant les biais développés par les outils d'IA du fait de l'emploi de jeux de données comportant eux-mêmes des défauts ;
- ➔ Le troisième agrège les réflexions autour des difficultés à identifier et suivre les failles des algorithmes qui pourraient résulter de la combinaison des deux premiers ensembles de problèmes.

Le traitement des questions éthiques liées l'IA de défense semble donc répondre, dans une certaine mesure, à la nécessité de limiter les freins qui pourraient s'exprimer quant à son développement et son déploiement. Il s'agit d'empêcher que l'accélération de l'intégration de l'IA dans l'appareil de défense ne soit entravée parce que ces questions, qui s'expriment déjà depuis de nombreuses années, suscitent des blocages.

Dans une certaine mesure, la position américaine vis-à-vis de l'IA de défense est donc caractérisée par des ambitions contradictoires. Alors que l'objectif premier est d'intégrer rapidement un premier jeu d'applications pour nourrir la dynamique, évaluer plus finement les potentiels des différentes technologies, perfectionner les procédures, former un grand nombre de personnels, disposer de cas pratiques permettant de réaliser du RETEX..., la nécessité de

⁵⁶ Pour une présentation synthétique de la TOS, voir Gros (P.), « La Third Offset Strategy américaine », *Défense&Industries*, 7, juin 2016, pp. 17-19.

⁵⁷ Klare (M.), « Pentagon Invests in AI, Issues Principles », *Arms Control Today*, May 2020.

traiter les questions éthiques est également affirmée. Or, si l'identification des interrogations éthiques que soulève l'IA de défense est déjà un travail complexe, la mise au point de réponses l'est encore plus. Il s'agit en effet d'une tâche qui demande du temps, nécessite des travaux scientifiques, se nourrit de débats juridiques et surtout ne peut être réalisée que de manière itérative – les projets menés à bien permettant de soulever de nouveaux problèmes.

► **Un rappel discret mais constant d'une IA au service des personnels**

Dans les documents institutionnels, apparaît à plusieurs reprises l'idée que l'Intelligence artificielle doit être « au service des personnels ». Cet objectif ne fait pas nécessairement l'objet de parties ou de paragraphe distinct(e)s dans ces références, mais se décline de plusieurs manières. Tout d'abord, il est affirmé – en particulier dans la Stratégie du DoD – que l'IA doit permettre d'autonomiser les personnels, et non de les remplacer. Il est même précisé qu'elle doit avant tout servir à les décharger des tâches fastidieuses. Ensuite, de nombreux développements dans les textes concernent la sécurité d'emploi pour les personnels, ainsi que les outils pour l'obtenir (en particulier la robustesse des méthodes de test, vérification et validation – voir ci-après – et la création de standards). De même, est affirmé le besoin de rendre les dispositifs compréhensibles par les opérateurs⁵⁸ – en particulier leurs comportements et les choix critiques qu'ils proposent aux humains. Dans certains documents du DoD, sont ainsi prévus des investissements pour les travaux sur « l'IA explicable »⁵⁹. Enfin, ces mêmes documents précisent que les personnels, notamment militaires, du DoD doivent être acteurs des réflexions et débats sur les questions éthiques. Alors que le traitement de ces problématiques est souvent confié à des conseils d'experts, certaines références du DoD, notamment la Stratégie, rappellent que ces débats doivent intégrer ceux qui vont utiliser les systèmes.

► **Un juridisme du traitement des questions éthiques**

L'étude des questions juridiques a été érigée en priorité institutionnelle dès la publication du premier document global de référence, le NDAA. Dans ce cadre, toutes les références du Département indiquent que le traitement des questions éthiques doit impliquer la conformité à des ensembles de droit, nationaux et internationaux, et même la création de nouvelles références juridiques.

► **Un fort lien questions éthiques/capacités de test-validation-vérification**

Dans les textes institutionnels produits par les différentes entités de la Défense américaine, un lien est très clairement fait entre les questions éthiques et les problèmes de test et d'évaluation des systèmes dotés d'IA. Dans ces documents, il est affirmé que les réponses à une partie des problèmes éthiques peuvent être obtenues en s'assurant que les domaines d'emploi des systèmes utilisant l'IA sont parfaitement connus. De même, la prévisibilité de ces systèmes doit pouvoir être assurée. Disposer de systèmes robustes, fiables et sécurisés est en soi

⁵⁸ Connaissance des méthodes ayant été employées pour développer et tester les systèmes, des domaines d'emploi, des technologies d'IA auxquels ils recourent...

⁵⁹ Entre autres efforts, la DARPA a notamment lancé, dès 2017, le programme « *Explainable AI* », qui cherche à fournir des modèles plus faciles à expliquer ou des voies pour le développement de modèles plus facilement explicables, afin d'aider les développeurs et les opérateurs à mieux comprendre les implications en termes de sûreté des systèmes d'IA.

Voir Fein (G.), « DARPA's XAI Seeks Explanations from Autonomous Systems », *Janes*, November 16, 2017.

une réponse à nombre de questions éthiques, qui justifie les efforts pour se doter de méthodes et outils de test, évaluation, vérification et validation. Les autorités militaires américaines, en particulier les responsables du JAIC, ont répété que le pays était en train de se positionner en *leader* mondial dans ce domaine⁶⁰. Dans la Stratégie du DoD, il est même indiqué que le Département doit être pionnier dans la conception de ces outils.

Parmi les méthodes disponibles pour atténuer les risques, les autorités américaines ont en effet décidé d'accentuer les efforts en matière de sécurité par le *design* et de test des systèmes eux-mêmes⁶¹. La DARPA a ainsi lancé un certain nombre de programmes de recherche dédiés à la fourniture d'assurances en matière de sécurité/sûreté. Le programme « *Assured Autonomy* », lancé en 2017, investigate par exemple les voies technologiques permettant de s'assurer que les systèmes apprenants puissent maintenir une sécurité/sûreté permanente et un fonctionnement correct bien qu'ils soient en train d'apprendre et s'adapter sans supervision humaine ou implication humaine lors de leurs entraînements.

► **La protection des valeurs et des intérêts américains comme finalité du traitement des questions éthiques**

Dans les documents institutionnels, le traitement des questions éthiques n'est pas seulement présenté comme une technique visant à favoriser l'acceptation de l'IA par les personnels et les populations et surtout à permettre d'assurer l'intégration rapide de systèmes au sein du DoD. Il est très clairement mis en rapport avec la protection des valeurs (notamment démocratiques) et des intérêts du pays. La dimension axiologique du discours sert notamment à polariser les positionnements face à l'IA : par le traitement des questions éthiques, avec leurs alliés, les États-Unis doivent pouvoir s'assurer que la conception et l'emploi des systèmes militaires recourant à l'IA se fassent en préservant certains ensembles de valeurs et corpus juridiques (protection de la vie privée, droits de l'Homme, discrimination civils/combattants...), alors que de nombreux pays ne les respectent pas. Dans le cadre de la compétition stratégique en matière d'IA, les autorités américaines ne cessent de répéter que certaines nations ne sont pas aussi « vertueuses » que les États-Unis (voir 3^{ème} Partie – Introduction).

Il est d'ailleurs intéressant de constater que les *think tanks* américains reprennent très largement cette opposition entre des États-Unis protecteurs des valeurs démocratiques et une Chine recourant à l'IA pour exprimer sa dimension autoritaire. Si les articles et études publiés sur les politiques développées par le DoD pour soutenir l'IA de défense peuvent être critiqués sur les documents américains, ils n'interrogent jamais cette opposition. Ils participent donc de sa diffusion.

Surtout, les textes institutionnels et les positions officielles rappellent que l'édification de cadres internationaux – notamment éthiques – pour le développement de l'IA doit permettre de protéger les intérêts américains. La notion d'« intérêt national » n'a jamais été précisément posée et demeure ambiguë, servant de registre de justification pour de nombreux types de

⁶⁰ Barnett (J.), « How the DoD is developing its AI ethics guidance », fedscoop, September 2, 2020.

⁶¹ Morgan (F. E.), Boudreaux (B.), Lohn (A. J.), Ashby (M.), Curriden (C.), Klima (K.), Grossman (D.), *Military Applications of Artificial Intelligence. Ethical Concerns in an Uncertain World*, Rand Corporation, 2020, p. 57.

décisions. Elle fait d'ailleurs l'objet de définitions parfois sensiblement différentes selon les pays⁶².

Dans les documents américains définissant la stratégie du pays à l'égard de l'Intelligence artificielle, en particulier l'*Initiative*, la composante économique des intérêts nationaux est présente. Il est ainsi clairement posé que l'engagement américain au niveau international a notamment pour but que les marchés internationaux demeurent ouverts aux solutions et produits des entreprises domestiques. Le soutien à la R&D fourni par l'État fédéral à ces mêmes firmes est conçu pour renforcer leur compétitivité, en particulier pour l'exportation. Les décisions américaines dans le domaine de la production des standards internationaux relèvent d'une logique équivalente : en certaines occasions (voir 3^{ème} Partie, § 4.), les autorités ont cherché à freiner des initiatives qui auraient pu créer des cadres trop rigides et surtout défavorables aux grands groupes américains.

Parmi la classification qu'il a produit des différents types d'intérêts nationaux, Robinson distingue, sur la scène internationale, les intérêts « identiques », « complémentaires » et « conflictuels »⁶³. Les premiers sont des sujets et objectifs qui sont communs à nombre d'États. La seconde catégorie se réfère à des intérêts, qui bien que n'étant pas strictement identiques, peuvent former le fondement d'un accord entre les pays sur des questions spécifiques. La dernière catégorie concerne des intérêts qui ne sont ni identiques, ni complémentaires.

Dans une assez large mesure, le discours américain sur les valeurs peut être vu comme un médium qui permettrait de faire du développement d'une IA éthique un intérêt « complémentaire », c'est-à-dire partagé par une communauté d'États « démocratiques », globalement respectueux de certains corpus de droits et désireux de les protéger. Mais il sert également à justifier la mise en place de coopérations scientifiques et technologiques bilatérales ou multilatérales et d'une action commune pour l'édification des standards internationaux qui est aussi conçue – de manière assez explicite – comme une méthode permettant aux entreprises américaines de dominer les marchés de l'IA.

Cette dichotomie des objectifs américains se retrouve également dans le domaine de la défense. La Stratégie du DoD insiste sur la mise en place de cadres pour s'assurer d'une IA responsable – en opposition notamment à la Chine. Elle a pour objectif de stimuler, en particulier sur le plan de la R&D, les alliés et les partenaires. Elle demande que les lignes directrices produites par le pays et les procédures de sécurité soient partagées. Mais ces éléments permettraient également aux entreprises américaines, déjà avancées dans ces domaines, d'imposer certaines normes à leurs concurrentes et ainsi de disposer d'un avantage sur elles.

⁶² Il convient toutefois de noter que les définitions officielles intègrent quasiment toutes l'auto-préservation (la capacité à demeurer légalement un véritable État), l'indépendance/souveraineté, le bien-être économique et la sécurité militaire.

⁶³ Robinson (T. W.), « National Interest », in Rosenau (J. N.), ed., *International Politics and Foreign Policy. A reader in research and theory*, New York: Free Press of Glencoe, Inc., 1961, pp. 184-85.

Structures internationales traitant des questions éthiques

De nombreuses structures internationales ont mis à leur agenda le traitement des questions liées au développement de l'Intelligence artificielle. Certaines des initiatives lancées partent notamment du constat que la pénétration progressive de l'Intelligence artificielle dans l'ensemble des activités sociales et socioprofessionnelles aura des conséquences sur les droits individuels et collectifs des personnes et sur leur bien-être. Par ailleurs, la capacité à doter progressivement les machines d'une autonomie décisionnelle de plus en plus grande oblige à réfléchir sur le rapport que les hommes vont entretenir avec ces systèmes.

Les applications militaires ont également été mises à l'agenda de certaines organisations internationales. Pour quelques-unes d'entre elles, ce choix a été sous-tendu par les actions développées par des acteurs issus de la société civile et de communautés scientifiques, qui ont dénoncé les risques liés au développement des systèmes d'armes létaux autonomes.

1. Mouvements de contestation des SALA

Depuis la fin des années 2000, plusieurs formes d'opposition aux drones armés, puis aux systèmes d'armes létaux autonomes, se sont structurées⁶⁴. Les critiques rassemblent notamment des organisations issues de la société civile (en particulier des ONG pacifistes), des intellectuels, des chercheurs et des cadres dirigeants d'entreprises de haute technologie. Ces différents acteurs ont lancé plusieurs campagnes internationales. Deux mouvements principaux peuvent être schématiquement isolés :

▶ **Une coalition d'ONG**

Une coalition d'une cinquantaine d'organisations non-gouvernementales s'est créée, coordonnée par *Human Rights Watch*.

▶ **Un collectif d'intellectuels et de chercheurs**

L'*International Committee for Robot Arms Control* (ICRAC) a été créé par 4 universitaires de renom (Noel Sharkey, Robert Sparrow, Jürgen Altmann et Peter M. Asaro), qui travaillent tous sur la problématique des robots. Il a été doté d'un statut d'ONG.

⁶⁴ La problématique la plus controversée concernant les SALA est la possibilité que les systèmes puissent prendre la décision de mettre fin, de manière autonome, à la vie d'un être humain et la mettre à exécution. Cette caractéristique confère à ces dispositifs une nature très différente de celle des drones. S'il existe bien une filiation entre les mouvements d'opposition aux drones armés (en particulier aux exécutions extrajudiciaires qu'ils ont permis de pratiquer) et aux SALA, ils ne sont pas confondus.

Ces deux mouvements ont lancé des campagnes contre les SALA à partir des années 2009-2010. Ils regroupent donc des acteurs issus d'horizons différents, qui n'ont pas tout à fait les mêmes modes de fonctionnement et répertoires d'actions, ni des objectifs complètement identiques. Ces détracteurs proposent toutefois des critiques assez proches.

Les appels et argumentaires de ces deux mouvements, ainsi que ceux d'acteurs intervenant de manière plus individuelle, sont diffusés aux niveaux global, régional, national et même local. Cette diffusion s'opère à la fois par les individus, mais également au travers de collectifs d'experts, d'ONG, d'entreprises de haute technologie, de divisions internes d'organisations internationales globales ou régionales et par l'action de certains gouvernements.

1.1. Principales critiques à l'égard des SALA

Certains arguments utilisés par les « anti-SALA » sont de nature morale et éthique. Pour beaucoup d'entre eux, le développement de l'autonomie reviendra à donner la capacité aux machines de décider de la vie ou de la mort des êtres humains. Certains vont même plus loin en dénonçant le risque que les machines puissent se rebeller contre l'homme. Ils recourent notamment à des œuvres littéraires et cinématographiques pour illustrer leurs propos (en particulier la série de films *Terminator*). Ils affirment la crédibilité de ces œuvres en rappelant que les capacités des machines à apprendre, en particulier de manière non-supervisée, et à évoluer par elles-mêmes font l'objet de nombreux travaux scientifiques. Ils considèrent que la conscience de soi des robots pourrait notamment émerger, indiquant que certaines composantes de cette conscience existent déjà. Depuis une vingtaine d'années environ, de nombreux spécialistes de robotique et d'IA (Hans Moravec, Kevin Warwick, Bill Joy, Hugo de Garis...) ont ainsi publiquement abordé la possibilité d'une « domination » des machines sur certaines existences humaines, voire sur les sociétés entières. D'autres ont mis l'accent sur la nécessité de concevoir dès l'origine une Intelligence artificielle bienveillante à l'égard des hommes.

Les réflexions proposées sont également de nature juridique. Il s'agit de l'impossibilité pour les machines d'être capables d'intégrer et de traduire en actions certains des principes qui fondent le droit international humanitaire s'appliquant lors des conflits. Dans ce domaine, les opposants affirment en effet que les systèmes autonomes vont notamment demeurer très longtemps incapables de véritablement discriminer les cibles légitimes et illégitimes. Dès lors, la machine risque de provoquer des dégâts, des blessures et des morts illicites. Or, avec l'acquisition de l'autonomie décisionnelle, se posera le problème de l'attribution de la responsabilité.

Pour les acteurs critiquant la conception et l'emploi des SALA, ces deux principaux problèmes vont nécessairement émerger car l'homme finira obligatoirement par être sorti de la boucle décisionnelle⁶⁵. Ils reprennent ainsi l'un des principaux motifs qui justifie, d'un point de vue opérationnel, l'utilisation des drones armés par les militaires (l'accélération de la boucle OODA) et poussent la logique jusqu'à l'une de ses limites. Ils affirment que l'obligation d'un homme confirmant les cibles et autorisant les frappes va devenir un désavantage, annulant

⁶⁵ Certains intervenants aux débats n'hésitent pas à affirmer que l'application du droit des conflits armés a pour conséquence l'impossibilité d'employer les systèmes autonomes pour opérer des tirs si un opérateur humain n'est pas « dans la boucle ».

pour partie la rapidité et la plus grande efficacité obtenue grâce à l'emploi des systèmes autonomes. Dans une certaine mesure, l'opérateur humain risque de représenter, à plus ou moins brève échéance, le « maillon faible » de la boucle⁶⁶. Par recherche d'efficacité et pour accélérer encore le tempo opérationnel, il sera donc très probablement éliminé des processus.

Certains arguments sont de nature plus politique. Les opposants aux SALA affirment que l'utilisation de machines à la place de combattants humains est susceptible de provoquer à terme une certaine « banalisation » des conflits. La moindre exposition des personnels militaires au danger, notamment par leur remplacement progressif dans les espaces où s'appliquent les feux, risquerait de faciliter la décision politique d'employer la force armée : les gouvernants qui choisiraient l'engagement de systèmes non habités ne prendraient en effet plus le risque de pertes humaines, toujours difficiles à accepter par les opinions publiques. Certains critiques vont même plus loin en indiquant que des conflits armés sans coût humain pourraient éroder le contrôle politique qui s'opère normalement en démocratie sur les activités militaires – les autorités civiles étant moins intéressées par les opérations et moins incitées à demander des comptes à la hiérarchie militaire.

1.2. Modes d'action et demandes

Les membres de l'ICRAC, pour la plupart des universitaires, ont une approche relativement académique de la thématique. Leurs positions sont d'ailleurs parfois relayées dans des revues ou des sites internet scientifiques ou de vulgarisation scientifique. S'ils n'hésitent pas à interpellier publiquement les opinions publiques, les autorités nationales et les organisations internationales sur la question, ils participent également aux controverses scientifiques sur les différentes thématiques qui sous-tendent la problématique de l'autonomie (par exemple lors de conférences scientifiques internationales). Du fait de leurs réelles compétences scientifiques, leur légitimité et le poids de leurs prises de position sont significatifs.

En juillet 2015, ces chercheurs ont toutefois cherché à étendre leur mouvement et a rassemblé au dehors de la communauté scientifique. Ils ont publié une lettre ouverte sur les dangers de l'autonomie des machines, à l'occasion de l'*International Joint Conferences on Artificial Intelligence* (IJCAI) qui a lieu à Buenos Aires. Cette lettre ouverte est en quelque sorte devenue une pétition puisqu'elle a été signée par 3 000 personnalités incluant des intellectuels et des figures des industries *Hi Tech* (comme Elon Musk, PDG de *Tesla Motors*, ou Steve Wozniak, cofondateur d'*Apple*)⁶⁷.

Les ONG ont, bien évidemment, une dimension beaucoup plus militante, fondée sur l'organisation de manifestations, la rédaction de lettres ouvertes aux responsables politiques, le lancement de pétitions... Depuis la fin des années 2000, certaines d'entre elles ont notamment

⁶⁶ Featherstone (S.), « The Coming Robot Army/Introducing America's Future Fighting Machines », *Harper's Magazine*, February 2007, pp. 43-49.

⁶⁷ Cette lettre ouverte a notamment été soutenue par le *Future of Life Institute*. En 2017, celui-ci a publié, avec une partie des premiers signataires, une seconde lettre, dont le propos était beaucoup plus centré sur les SALA.

publié des rapports sur le problème des SALA⁶⁸. Une campagne internationale visant à l'interdiction des « robots tueurs » (campagne « *Stop Killer Robots* ») a été lancée en avril 2013. Coordonnée par l'ONG américaine *Human Rights Watch*, elle est essentiellement menée dans une vingtaine de pays.

Les deux mouvements se sont rapprochés avant même le lancement de la campagne, qui a reçu le soutien d'autres acteurs. En mai 2014, 20 lauréats du prix Nobel ont publié une déclaration conjointe dans laquelle ils appelaient à soutenir cette initiative. De même, en novembre 2014, une déclaration inter-religieuse est venue également promouvoir la campagne. Elle a été signée par 70 *leaders* religieux.

La principale demande de ces différents acteurs est celle d'un moratoire sur le développement scientifique et technologique des systèmes autonomes. La campagne internationale coordonnée par l'ONG américaine *Human Rights Watch* réclame ainsi que tous les pays disposant de programmes de recherche et surtout de développement sur l'autonomie des systèmes armés les suspendent. Cette « pause » doit notamment permettre d'éviter que la conception de dispositifs véritablement autonomes devienne un fait accompli, c'est-à-dire soit opérée dans un avenir plus ou moins proche sans même que des réflexions poussées les aient concernés et qu'un cadre juridique adapté – notamment en droit international – n'ait été conçu⁶⁹.

Avant même que ce cadre juridique spécifique ne soit créé, la campagne enjoint les autorités nationales d'affirmer (ou de réaffirmer) leur engagement à respecter le droit international humanitaire et le droit international relatif aux droits de l'Homme lors des emplois de « robots armés ». De même, elle leur demande de mettre en place des procédures permettant d'assurer la conformité de toutes les étapes de la conception de ces systèmes avec ces corpus juridiques.

La campagne les incite aussi à s'engager à être les plus transparents possibles sur les procédures internes d'évaluation de ces équipements (notamment sur les métriques et les tests employés pour mesurer les aptitudes des systèmes et les assurances en termes de fiabilité, sécurité et sûreté). Enfin, elle leur demande de participer aux débats internationaux sur la question des systèmes létaux autonomes (notamment dans l'objectif d'échanger sur les meilleures pratiques dans ce domaine et pour promouvoir la mise en place de garde-fous partagés).

Les acteurs les plus opposés aux systèmes autonomes proposent même de renforcer le cadre juridique en créant un traité d'interdiction, sur le modèle de ceux qui ont été mis en place pour les mines antipersonnel et les armes à sous-munitions. Pour certains d'entre eux, il s'agit donc de développer un dispositif d'*arms control* préventif⁷⁰. Habituellement, la maîtrise des

⁶⁸ Par exemple, le rapport commun de *Human Rights Watch* et de la *Harvard Law School's International Human Rights Clinic*, intitulé *Losing Humanity: The Case Against Killer Robots*, paru en novembre 2012.

⁶⁹ Ces acteurs constatent notamment que nombre de développements civils sont duaux et qu'ils facilitent donc la conception de SALA.

Dès le lancement de leurs actions, ils ont affirmé que les réflexions sur les SALA devaient être produites immédiatement, même si l'autonomie et l'intelligence des systèmes demeurent encore très limitées (voir par exemple Bode (I.), Huelss (H.), « Why 'Stupid' Machines Matter: Autonomous Weapons and Shifting Norms », *Bulletin of the Atomic Scientists*, 12 October 2017).

⁷⁰ Sur ces éléments, voir Bouterin (G.), « Les drones, futurs objets d'*arms control* ? », dans CESA, *Les drones aériens : passé, présent et avenir. Approche globale*, Paris, La Documentation française, 2013, pp. 551-573.

armements est mise en place alors même que les armements problématiques sont en dotation dans les armées et ont même déjà été employés lors de conflits. Il vise dès lors à limiter les quantités détenues par chaque pays ou à empêcher les évolutions qualitatives des armes. La solution actuellement prônée est de réguler, voire d'interdire, les systèmes armés autonomes alors même qu'ils n'existent pas encore véritablement et n'ont donc pas été introduits dans les arsenaux⁷¹.

Plus précisément, certains organismes animateurs de la campagne internationale demandent que des discussions soient menées autour de plusieurs thèmes :

- L'interdiction de tout nouveau développement, acquisition, déploiement et emploi de systèmes d'armes autonomes (c'est-à-dire un moratoire qui permettrait juste de continuer à utiliser l'existant et d'en tirer des enseignements) ;
- La limitation de la portée et des armements pouvant être emportés par les systèmes dotés d'autonomie, mais dont le fonctionnement continuerait de reposer sur le concept d'« un homme dans la boucle » ;
- L'interdiction du déploiement dans des postures menaçant d'autres États ;
- L'interdiction de doter les systèmes inhabités d'armes nucléaires ;
- L'interdiction du développement, du déploiement et de l'emploi de systèmes d'armes autonomes dans l'espace.

Les demandes des différents acteurs ne sont donc pas nécessairement les mêmes. Elles sont même parfois incompatibles. Si certains réclament un traité d'abolition des SALA, d'autres militent pour un moratoire sur les recherches scientifiques – action plus limitée. L'ONG *Landmine Action* était pour sa part plus précise et demandait une interdiction uniquement centrée sur les systèmes disposant de capacités de ciblage autonomes (mais acceptant par exemple les développements technologiques en matière de navigation autonome).

2. Principales organisations internationales prenant en charge des débats sur les problématiques éthiques liées à l'IA

2.1. Nations Unies

Un certain nombre d'institutions des Nations Unies et certains de ses satellites ont abordé, ces dernières années, la thématique de l'Intelligence artificielle, en particulier sous l'angle des problèmes éthiques et juridiques que son développement soulève. Plus spécifiquement, la Convention sur certaines armes conventionnelles est la principale enceinte au sein de l'« archipel » des Nations Unies à avoir pris en charge les discussions sur les applications militaires de l'IA, en particulier la problématique des SALA.

⁷¹ Voir le texte d'un des principaux promoteurs de cette solution : Altmann (J.), « Preventive Arms Control for Uninhabited Military Vehicles », in Capurro (R.), Nagenborg (M.), eds., *Ethics and Robotics*, Heidelberg, AKA Verlag, 2009.

2.1.1. UNESCO

Dès 2017, La Commission mondiale d'éthique des connaissances scientifiques et des technologies (COMEST) de l'UNESCO, organe consultatif et forum de réflexion mis en place en 1998, a publié un rapport portant sur les questions éthiques liées au développement des différentes formes de robotique⁷². Ce document traitait notamment des défis posés, dans le domaine éthique, par l'emploi de plus en plus diversifié des robots dans la société. Il fournissait un cadre de travail pour penser les questions liées au développement de ces technologies, ainsi que des principes et des valeurs applicables au recours aux robots.

L'UNESCO a par ailleurs organisé, dans ses locaux de Paris, un débat public en janvier 2019⁷³. Il s'agissait d'ouvrir une discussion autour des aspects éthiques du développement des nouvelles technologies de rupture, dont l'Intelligence artificielle. Le débat a été articulé autour d'experts (scientifiques, juristes, ingénieurs, philosophes, sociologues et représentants d'associations).

En novembre 2019, lors de la 40^{ème} session de l'organisation, la Conférence générale a par ailleurs décidé d'initier un processus d'élaboration d'un document normatif international concernant l'éthique de l'Intelligence artificielle. Ce cycle est prévu pour durer 2 ans. Initiative inclusive et multidisciplinaire, il repose sur l'organisation de consultations avec de très nombreuses parties prenantes. Suite à ces consultations, un premier texte, proposant des recommandations, a été publié en septembre 2020⁷⁴. Il s'agit d'un projet, pour l'amélioration duquel des discussions en ligne ont été lancées.

2.1.2. Union internationale des télécommunications (UIT)

En 2017, l'UIT a lancé un programme intitulé « L'intelligence artificielle au service du bien social ». Celui-ci s'est notamment manifesté par l'organisation d'un Sommet mondial. En tant qu'institution spécialisée des Nations Unies, l'UIT dispose en effet de la légitimité pour lancer des réflexions sur l'IA. Surtout, elle rassemble des représentants des gouvernements, des entreprises, du monde académique (en particulier des universités) et de mouvements issus de la société civile.

Ce sommet a été par la suite reconduit sur une base annuelle. Il constitue désormais l'un des forums privilégiés par les Nations Unies pour organiser un dialogue international large concernant l'Intelligence artificielle. Sa fonction est d'accélérer la conception et le déploiement d'applications pratiques de l'IA dans les domaines liés aux Objectifs de développement durable définis par les Nations Unies.

En 2019, il a notamment permis de rassembler, entre autres participants, des représentants d'une trentaine d'institutions des Nations Unies et d'organisations internationales (comme

⁷² UNESCO, World Commission on the Ethics of Scientific Knowledge and Technology, *Report of COMEST on Robotics Ethics*, SHS/YES/COMEST-10/17/2 REV, 14 September 2017.

⁷³ Événement intitulé « Debate on Ethics of New Technologies and Artificial Intelligence. Tech Futures: Hope or Fear? ».

⁷⁴ United Nations Educational, Scientific and Cultural Organization, Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, *First Draft of the Recommendation on the Ethics of Artificial Intelligence*, SHS/BIO/AHEG-AI/2020/4 REV.2, September 2020.

l'UE). Cet événement a notamment servi à présenter les stratégies adoptées par certains États, secteurs économiques, organisations internationales..., qui permettent d'assurer que les technologies de l'Intelligence artificielle soient élaborées de manière sûre et inclusive, en garantissant qu'elles n'accentuent pas les inégalités (par un accès équitable aux avantages qu'elles offrent). Il a aussi été utilisé pour concevoir des propositions de projets innovants sur le thème global de l'IA au service du bien social⁷⁵.

L'UIT a par ailleurs décidé de mettre à disposition certains outils pour aider les acteurs qui veulent réfléchir aux impacts de l'IA. Un répertoire a ainsi été constitué, que toute personne travaillant dans le domaine de l'Intelligence artificielle peut nourrir. Il s'agit d'une base de données, proposant des informations sur l'utilisation de l'IA au profit du bien social. Il recense notamment les projets, travaux de recherche, groupes de réflexion et initiatives lancés par les organisations internationales.

Autre action notable, l'UIT rassemble régulièrement des responsables d'organismes de réglementation des TIC venus du monde entier pour échanger sur l'Intelligence artificielle, et notamment sur les questions réglementaires qui sont associées à ses différents emplois. Ces réunions permettent également d'examiner des points relatifs à la gouvernance.

En matière de normalisation, un Groupe spécialisé a été créé (Groupe spécialisé sur l'apprentissage automatique pour les réseaux futurs), dont la mission est d'examiner comment la normalisation technique peut appuyer les nouvelles applications d'apprentissage automatique dans des domaines tels que l'analyse des mégadonnées, la sécurité et la protection des données⁷⁶. Il s'agit notamment de travailler sur ces thématiques alors que le déploiement de la 5G a commencé. Ce groupe élabore des projets de spécifications afin que les réseaux de TIC et leurs composants soient capables d'adapter de manière autonome leur comportement pour respecter les aspects liés à l'éthique, à l'efficacité et à la sécurité, tout en garantissant une expérience optimale pour l'utilisateur.

Au lendemain de l'édition de 2019 du Sommet mondial sur l'intelligence artificielle au service du bien social, l'appel en faveur d'une normalisation accrue pour les applications de santé a été suivi par la création d'un autre Groupe spécialisé (Groupe spécialisé sur l'intelligence artificielle au service de la santé). Cet organisme a notamment pour mission de définir un cadre de référence applicable à l'évaluation des algorithmes d'Intelligence artificielle utilisés pour des applications de soins.

2.1.3. UN Convention on Certain Conventional Weapons (CCW)

Différentes institutions internationales ont décidé de se positionner sur la thématique des systèmes d'armes autonomes et de l'insérer à leur agenda, en partie pour répondre aux demandes des mouvements anti-SALA (voir ci-avant). Dès 2013, le rapporteur spécial de l'ONU

⁷⁵ Idées de projets dans les domaines de l'amélioration des soins de santé, du suivi de l'agriculture et de la biodiversité...

⁷⁶ Les groupes spécialisés de l'UIT-T sont des structures souples dont les cycles de travaux sont généralement brefs (en général 1 à 2 ans). Ils servent à accélérer les études menées dans des domaines qui revêtent un intérêt stratégique croissant pour les membres de l'UIT. Inclusifs, ils sont ouverts à toute partie intéressée. L'une de leurs principales fonctions est de poser les jalons des travaux de normalisation connexes des Commissions d'études de l'UIT-T.

sur les exécutions extrajudiciaires préconisait ainsi, dans un rapport présenté au Conseil des droits de l'homme, un moratoire sur les robots létaux autonomes⁷⁷. Le Secrétaire général des Nations Unies et le chef du Bureau des affaires du désarmement des Nations Unies ont pour leur part exprimé, à plusieurs reprises, leur volonté que le sujet des SALA soit attentivement étudié.

En novembre 2013, plus d'une centaine d'États parties à la Convention sur certaines armes conventionnelles ont accepté de traiter la question des SALA dans le cadre de la Convention⁷⁸. Celle-ci est un forum relativement consensuel car elle dispose d'une certaine flexibilité (en particulier quant aux thématiques qu'elle traite) et que les débats qui y sont menés intègrent à la fois l'application des principes humanitaires mais également les nécessités militaires. Par ailleurs, les principaux pays développant des capacités en IA à cette date étaient des Hautes parties contractantes à la Convention (Canada, Chine, France, Allemagne, Inde, Israël, Japon, République de Corée, Fédération de Russie, Royaume-Uni et États-Unis)⁷⁹.

Du 13 au 16 mai 2014, s'est ainsi tenue à Genève une « Réunion informelle d'experts sur les Systèmes d'Armes Létaux Autonomes (SALA) ». Elle a permis de rassembler des représentants de 87 pays, ainsi que des délégués de la coalition d'ONG « Stop Killer Robots ». En soi, ce rattachement à la Convention peut d'ailleurs être considéré comme une victoire pour les « anti-robots tueurs » puisqu'elle constitue un cadre multilatéral de limitation et d'interdiction de certains armements. La rapidité avec laquelle la problématique a été inscrite à l'ordre du jour d'une institution pourtant réputée pour la lenteur de ses débats est également intéressante à relever. Au cours de la première réunion, une cinquantaine de pays ont rendu publiques des positions officielles concernant la question.

La première réunion a été prolongée par une seconde rencontre en novembre 2014. Une troisième s'est déroulée en avril 2016⁸⁰. Le fait que les rencontres sur le sujet s'enchaînent et donc que la problématique reste à l'agenda international peut être également considéré comme une victoire pour les opposants aux « robots » armés autonomes. Surtout, l'ONU a décidé de créer un Groupe d'experts gouvernementaux (*Group of Governmental Experts – GGE*) sur les

⁷⁷ Christof Heyns, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, A/HRC/23/47*, Geneva, 2013.

⁷⁸ La Convention de 1980 sur l'interdiction de certaines armes classiques (ou « Convention de 1980 sur l'interdiction ou la limitation de l'emploi de certaines armes classiques qui peuvent être considérées comme produisant des effets traumatiques excessifs ou comme frappant sans discernement ») compte cinq parties ou « protocoles ». Il s'agit d'un accord multilatéral interdisant ou limitant l'emploi de certaines armes classiques (pièges ayant l'apparence d'objets inoffensifs, mines...), ouvert à la signature le 10 avril 1981 et entré en vigueur le 2 décembre 1983. Les racines de la Convention sont les principes clés du droit international humanitaire (proportionnalité, distinction civils/combattants...).

⁷⁹ Le traitement de la question des SALA par la Convention soulevait cependant des problèmes. Tout d'abord, la continuité du financement des actions de concertation n'était pas assurée. Un certain nombre de Hautes parties contractantes étaient en effet à cette époque en retard de leurs contributions au budget. Par ailleurs, pour que les débats puissent être les plus riches possibles, il était nécessaire d'intégrer les industries et les développeurs technologiques aux discussions.

⁸⁰ Ces premières rencontres ont notamment servi à réaliser une revue trans-domaine de la littérature disponible, afin de rendre compte de toutes les dimensions du problème (éthique, humanitaire, militaire, juridique, technologique et commercial).

SALA pour rendre plus formelles les discussions⁸¹. Ce choix a été validé par les États-parties en décembre 2016. Cet organe se réunit chaque année dans le cadre de la Convention sur certaines armes conventionnelles et même de la Conférence sur le Désarmement. De la sorte, le traitement de la thématique a été pérennisé et l'organisme qui centralise le processus – le groupe d'experts – a été institutionnalisé : ce traitement est passé de réunions informelles à des rencontres organisées dans des cadres pérennes sur la base d'un mandat de travail.

En dehors des experts nationaux et des représentants des deux principaux mouvements militant contre les SALA (voir ci-avant, 2^{ème} Partie, § 1.), différentes organisations ont contribué à nourrir les travaux. Le Comité international de la Croix rouge avait déjà travaillé sur la thématique et a pu proposer des réflexions. Des *think tanks*, comme le *Stockholm International Peace Research Institute* et *Chatham House*, ont également lancé des programmes dédiés – tout comme quelques structures de recherche académiques (en particulier la *Harvard Law School*). De la sorte, l'expertise mobilisée a pu être nombreuse et d'origines relativement diverses. Surtout, l'*United Nations Institute for Disarmament Research* (UNIDIR), en tant que centre de recherche interne autonome des Nations Unies travaillant sur les questions de désarmement, a fourni une contribution substantielle en développant un document d'amorce et d'autres matériaux, pour les rencontres, à destination des négociateurs et des chercheurs.

Progressivement, de plus en plus de parties contractantes se sont, officieusement ou officiellement, prononcées pour la création d'un instrument légalement contraignant concernant les systèmes d'armes totalement autonomes. À l'occasion de la première réunion du GGE (novembre 2017), clôturée par une déclaration indiquant que les participants considéraient que la CCW était un cadre approprié pour traiter la thématique des SALA, 22 pays ont ainsi appelé à une interdiction légale préemptive des SALA. Ce groupe n'intégrait cependant aucune nation majeure du point de vue technologique⁸².

En avril 2018, un jeu de 10 possibles principes a été constitué. Il s'agissait d'un ensemble de règles qui pourrait être adopté par les Hautes parties contractantes :

- ➔ Applicabilité du droit international humanitaire aux SALA ;
- ➔ Non-délégation de la responsabilité humaine ;
- ➔ Mise en responsabilité pour l'emploi de la force en accord avec le droit international ;
- ➔ Processus d'examen des systèmes d'armes avant leur déploiement ;
- ➔ Intégration de garanties pour la sécurité physique, la non-prolifération et la cybersécurité ;

⁸¹ Il existe par ailleurs un *Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, établi par l'Assemblée générale en 2004. Dans le cadre de son mandat, il est également amené à traiter d'IA.

⁸² Algérie, Argentine, Bolivie, Brésil, Chili, Costa Rica, Cuba, Équateur, Égypte, Ghana, Guatemala, Saint-Siège, Irak, Mexique, Nicaragua, Pakistan, Panama, Pérou, Palestine, Ouganda, Venezuela et Zimbabwe.

Au cours de la Conférence d'examen de la CCW de 2016, la Chine avait rendu publique une première position officielle sur les systèmes d'armes létaux autonomes. Celle-ci précisait que le pays soutient le développement d'un protocole légalement contraignant sur les questions liées à l'utilisation des SALA – protocole similaire à celui de l'interdiction des armes laser aveuglantes. La Chine n'avait cependant pas pris soin à cette époque de préciser ce qu'elle entendait par SALA.

- ➔ Évaluation et atténuation des risques durant la phase de développement des technologies ;
- ➔ Examen de la conformité au droit international humanitaire de l'emploi des technologies émergentes dans le domaine des SALA ;
- ➔ Absence de préjudice à la R&D et aux emplois civils ;
- ➔ Besoin d'adopter une perspective non-anthropomorphique pour l'IA ;
- ➔ Caractère approprié de la Convention sur certaines armes classiques comme cadre pour traiter de la question des SALA⁸³.

En novembre 2019, après des négociations autour de cet ensemble, les États parties ont approuvé, lors d'une réunion, 12 principes directeurs. Ces principes affirment notamment que le droit international humanitaire a vocation à s'appliquer à tous les systèmes d'armes, y compris ceux dotés d'autonomie, et que l'emploi de ces systèmes doit toujours dépendre d'une responsabilité humaine. Ils traitent de la nécessité d'une interaction entre l'humain et la machine, dont la nature et l'étendue n'ont cependant pas été précisées, pour assurer le respect du droit international humanitaire lorsque ces armes sont utilisées. Ces principes rappellent aussi que les États doivent, conformément au premier protocole additionnel aux Conventions de Genève, examiner la licéité des nouvelles armes qu'ils développent.

Si ces principes ont bien été approuvés, il convient toutefois de préciser qu'en mars 2019, lors d'une réunion du Groupe d'experts gouvernementaux, de nombreux pays (dont la France, les États-Unis, l'Australie, la Chine et la Russie) se sont prononcés contre l'adoption d'un instrument international juridiquement contraignant pour encadrer le développement et l'utilisation des armes autonomes.

2.2. Union européenne

Pendant plusieurs années, l'Union européenne a travaillé sur différents aspects de l'Intelligence artificielle sans les lier. Elle n'avait pas développé de réflexion spécifique et globale sur l'IA, qui pourtant apparaissait dans des documents institutionnels assez divers.

À partir de 2016, différents documents ont appelé à accentuer les efforts de réflexion européens sur l'IA⁸⁴. En janvier 2017, la Commission juridique du Parlement européen a notamment adopté un texte non-contraignant contenant des recommandations pour la *Commission on Civil Law Rules on Robotics*⁸⁵. Ce document intégrait une demande pour que la Commission européenne étudie la possibilité de créer une Agence européenne dédiée et qu'elle explore,

⁸³ Alors même que certains « anti »-SALA, en particulier les organismes rassemblés dans la campagne « *Stop Killer Robots* », critiquent le rythme trop lent des discussions en son sein.

⁸⁴ Dont la *Stratégie globale pour la politique étrangère et de sécurité de l'Union européenne*.

⁸⁵ European Parliament, Committee on Legal Affairs, Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), 27.1.2017.

En janvier 2015, la Commission juridique du Parlement européen a décidé de créer un groupe de travail dédié aux questions juridiques suscitées par le développement de la robotique et de l'IA.

analyse et considère les implications de la création d'un statut légal pour les robots⁸⁶. La Commission juridique du Parlement donnait aussi pour mission à la Commission de se doter de l'expertise technique, juridique et sur les questions éthiques nécessaire pour analyser le développement d'une société largement fondée sur l'IA.

De même, en mai 2017, le Comité économique et social européen publiait son approche de l'IA⁸⁷. Il s'était autosaisi du dossier. Dans son Avis d'initiative, il appelait à la création d'un code éthique devant encadrer le développement, les applications et les utilisations de l'Intelligence artificielle.

En octobre 2017, le Conseil européen a estimé qu'il existait un caractère d'urgence au fait d'étudier certaines tendances technologiques émergentes, notamment le développement de l'IA. Il s'est affirmé préoccupé par le fossé qui semblait grandir entre l'Europe d'un côté et les États-Unis et la Chine de l'autre. Il a ainsi invité la Commission à aller plus loin dans ses démarches et à produire une véritable approche européenne au début de l'année 2018. Afin d'analyser les questions éthiques liées à l'IA, le Commissaire européen à la Recherche, à l'Innovation et à la Science, Carlos Moedas, a ainsi créé un groupe d'experts chargé de rédiger un rapport (*European High-Level Group on Artificial Intelligence* – voir ci-après).

2.2.1. Documents majeurs

A. Communication de la Commission européenne (avril 2018)

En avril 2018, la Commission publiait, comme demandé par le Conseil européen, sa Communication « *Artificial intelligence for Europe* »⁸⁸. Ce document, d'une vingtaine de pages, a permis de fixer une approche européenne de l'IA. Il a servi à créer un cadrage global, pour que soit rédigée une véritable stratégie, ainsi que pour la conception du plan d'investissement qui doit l'accompagner.

Par la publication de cette Communication, la Commission européenne a cherché à indiquer qu'elle considère l'Intelligence artificielle comme une technologie stratégique et qu'elle fait du soutien à son développement une priorité. Le document sert notamment à appeler les États-membres à mettre en place une coordination de leurs stratégies nationales.

La Communication met essentiellement l'accent sur les efforts que l'Union européenne doit consentir en matière de recherche et pour la transformation de quelques secteurs socioéconomiques choisis (santé et transport notamment). Elle sert à informer des investissements

⁸⁶ Dans le document, le Parlement proposait aussi un code éthique pour les ingénieurs roboticiens.

⁸⁷ Comité économique et social européen, *L'intelligence artificielle – Les retombées de l'intelligence artificielle pour le marché unique (numérique), la production, la consommation, l'emploi et la société*, Avis, INT/806 – EESC-2016-05369-00-00-AC-TRA (NL) 1/14, 31 mai 2017.

⁸⁸ European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Artificial Intelligence for Europe*, COM(2018) 237 final, Brussels, 25.4.2018.

En droit européen, une Communication est un document qui ne crée aucune obligation juridique en lui-même. Il s'agit d'un « acte préparatoire » qui exprime la position de la Commission sur un sujet donné et présente ses propositions pour le traiter. Il se situe en amont du processus de décision.

que la Commission compte consentir pour le soutien de l'IA, afin de susciter la participation des États, des collectivités locales et des acteurs privés.

L'approche européenne de l'IA présentée dans le document est fondée sur 3 piliers :

- ➔ Favoriser les développements technologiques et industriels et encourager l'exploitation par les secteurs privés et publics ;
- ➔ Préparer les Européens aux changements socio-économiques engendrés par l'introduction de l'IA ;
- ➔ S'assurer qu'un cadre éthique et légal approprié soit créé.

Dans le cadre du troisième pilier, la Communication de la Commission souligne que l'Union doit impérativement protéger les données à caractère personnel, mais également la sécurité des données et rappelle la responsabilité du producteur. Pour s'assurer que ces exigences pourront être véritablement imposées, l'UE annonce qu'elle va recourir à son pouvoir de création des normes.

Dans la Communication, il est indiqué que la Commission compte produire, pour la fin de l'année 2018, un projet de lignes directrices éthiques sur l'IA, devant servir de support au travail du groupe d'experts de haut niveau dédié.

Certains observateurs ont vu dans la publication de cette Communication un moyen pour la Commission de marquer sa volonté de promouvoir une approche de l'IA respectueuse des valeurs fondamentales de l'Europe⁸⁹. Le document vise en effet à soutenir une IA « éthique » et « positive », c'est-à-dire dont le développement ne fait pas l'impasse sur le traitement des problèmes liés aux utilisations sensibles et respecte les droits fondamentaux européens (en particulier dans les domaines de la protection des données, de l'ouverture et de la transparence).

B. Déclaration commune de soutien à l'approche de l'UE (avril 2018)

Au cours de l'*EU Digital Day*, les États membres ont signé une déclaration de soutien à l'approche européenne présentée par la Commission.

La Déclaration met en avant plusieurs principes que les États-membres voudraient voir absolument traités par la Commission :

- ➔ L'accès aux données du secteur public ;
- ➔ L'atténuation des défis socio-économiques induits par l'emploi croissant des technologies reposant sur l'IA ;
- ➔ Le développement d'un cadre légal et éthique pour l'IA fondé sur les valeurs et les droits fondamentaux européens (et incluant la protection de la vie privée et des données à caractère personnel, ainsi que des principes de transparence et de responsabilité).

⁸⁹ Chrétien (J.), *Intelligence artificielle : bâtir la voie européenne*, Note stratégique, EU Digital Challenges, octobre 2018.

En application de l'ébauche de stratégie produite par la Commission, l'*European High-Level Group on Artificial Intelligence* a effectivement été créé en décembre 2018. Une liste préliminaire de garde-fous éthiques a été publiée par la Commission au même moment. Elle incluait notamment la prohibition de l'identification de la personne dont les données sont employées pour « nourrir » les systèmes d'IA, la transparence concernant les applications de l'IA et la préservation d'un contrôle humain sur les productions technologiques.

C. *Publication des Ethics Guidelines for Trustworthy Artificial Intelligence (avril 2019)*⁹⁰

Les Lignes directrices en matière d'éthique sont contenues dans un rapport d'une cinquantaine de pages. Bien que rédigé par l'*European High-Level Group on Artificial Intelligence*⁹¹, le document a été approuvé, et donc endossé, par la Commission.

Il s'agit d'un jeu de lignes directrices applicables au développement et à l'utilisation des systèmes recourant à l'Intelligence artificielle. Ces normes sont non-contraignantes : elles doivent être respectées volontairement, afin de servir à guider les concepteurs et les utilisateurs, mais également à rassurer les parties prenantes et les opinions nationales, de manière à renforcer la diffusion de systèmes sûrs et éthiques.

Le document précise qu'il est nécessaire que l'IA dispose de trois caractéristiques pour être « digne de confiance ». Elle doit être :

- Légale (c'est-à-dire respectueuse de l'ensemble des lois et réglementations qui peuvent s'appliquer à elle) ;
- Éthique (c'est-à-dire respectueuse des principes éthiques et des valeurs communes des membres de l'UE) ;
- Robuste (d'un point de vue technique et en prenant en compte son environnement social).

Les Lignes éthiques posent un ensemble de 7 conditions majeures pour que les systèmes recourant à l'Intelligence artificielle puissent être « dignes de confiance ».

▶ **Intervention humaine et supervision**

Pour le groupe d'experts de haut niveau, les systèmes d'Intelligence artificielle doivent « renforcer » les êtres humains, c'est-à-dire les rendre capables de prendre des décisions mieux informées tout en favorisant leurs droits fondamentaux. Des mécanismes de supervision (voire de surveillance) adaptés doivent être mis en œuvre, au travers d'approches conservant un homme dans la boucle, fondées sur un humain sur la boucle ou reposant sur un humain aux commandes.

⁹⁰ Groupe d'experts indépendants de haut niveau sur l'Intelligence artificielle constitué par la Commission européenne en juin 2018, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, avril 2019.

⁹¹ L'*European AI Alliance*, plate-forme réunissant de nombreuses parties prenantes d'États-membres différents soutenue par la Commission, a notamment fourni des inputs au Groupe de haut niveau.

▶ **Robustesse technologique et sécurité**

Pour les auteurs, les systèmes dotés d'Intelligence artificielle doivent être résilients et sûrs. Ils doivent intégrer un plan de rechange au cas où quelque chose d'anormal surviendrait. Ils doivent être aussi précis, fiables et reproductibles. De la sorte, le risque de survenue de dommages involontaires pourra être minimisé.

▶ **Vie privée et gouvernance des données**

Afin d'assurer un plein respect de la vie privée et de la protection des données, les experts considèrent que des mécanismes adaptés de gouvernance des données doivent être assurés. Ils doivent prendre en compte la qualité et l'intégrité des données et assurer que les accès sont justifiés (légitimes).

▶ **Transparence**

Pour le groupe, l'exploitation des données, le système et les modèles économiques doivent être transparents. Des mécanismes de traçabilité peuvent aider à respecter ces conditions. Surtout, les dispositifs recourant à l'Intelligence artificielle et leurs décisions doivent pouvoir être expliqués de manière adaptée aux parties prenantes concernées. Les humains ont besoin d'être avertis qu'ils sont en interaction avec un système intégrant de l'Intelligence artificielle et doivent être informés sur les capacités du système, mais également ses limites.

▶ **Diversité, non-discrimination et équité**

Des biais provoquant des discriminations peuvent survenir, générant d'éventuelles conséquences négatives multiples pour les individus (marginalisation d'un groupe social vulnérable, exacerbation de préjugés de personnes déjà victimes...). Les experts de haut niveau affirment donc que les systèmes utilisant de l'IA doivent encourager la diversité et être accessibles à tous (y compris aux individus handicapés). Tout au long du cycle de vie de ces dispositifs, les parties prenantes doivent être impliquées.

▶ **Bien-être sociétal et environnemental**

Pour les auteurs, les systèmes d'IA doivent bénéficier à tous les humains – ce qui inclut les générations futures. Il faut ainsi s'assurer qu'ils sont durables et protecteurs de l'environnement. De plus, ils doivent être capables de prendre en compte des conséquences indirectes sur l'environnement, les groupes sociaux et la société.

▶ **Mise en responsabilité**

Le rapport indique que des mécanismes doivent être créés pour s'assurer qu'une responsabilité des systèmes d'Intelligence artificielle et des conséquences de leur emploi existe. Une vérifiabilité doit exister, qui implique qu'une évaluation des algorithmes, des données et des procédures puisse être pratiquée, notamment lors des applications critiques. De même, une capacité de rectification doit toujours être présente.

Le document comprend des outils pour vérifier que ces conditions sont bien respectées. Il propose ainsi une liste permettant de réaliser une évaluation des dispositifs. Il a été demandé aux parties prenantes de tester cette liste à partir de juin 2019, afin de fournir des retours au groupe d'experts. Cette phase d'expérimentation s'est prolongée jusqu'à la fin de l'année 2019.

Un certain nombre de commentateurs ont critiqué le manque de clarté des Lignes directrices. Quelques-uns des membres du groupe d'experts ont même reconnu que ces lignes sont effectivement vagues et développées pour le court terme⁹². Leur création n'aurait pas intégré les risques de long terme. Ces rédacteurs regrettaient aussi que certaines « lignes rouges » (des principes éthiques non-négociables) présentes dans le projet de rapport n'aient finalement pas été reprises dans la version finale. L'une de ces limites concernait les SALA : il était en effet initialement indiqué que l'IA ne devait jamais être employée pour la création de systèmes létaux autonomes (à quelque niveau que ce soit). Des négociations ont amené des modifications : les catégories d'applications initialement prohibées n'ont plus été citées et simplement réunies sous l'expression « préoccupations majeures », sans qu'une interdiction formelle ne soit émise. De même, il a été noté le manque d'une hiérarchie entre les principes posés par les Lignes directrices. Cette priorisation aurait sans doute permis aux institutions européennes de concevoir plus facilement leurs politiques de soutien.

Plus largement, les Lignes éthiques ont un caractère non-contraignant. Or, même au niveau national, il existe très peu de réglementations visant à soutenir leur mise en œuvre. De même, il existe très peu d'incitations à y adhérer.

Le principe qui constitue le cœur de la politique européenne en matière d'Intelligence artificielle est celui d'une approche centrée sur l'humain. Il a notamment guidé la rédaction des Lignes directrices en matière éthique. Il a été défini par les experts de haut niveau :

« L'approche de l'IA centrée sur l'humain s'efforce de garantir que les valeurs humaines soient un élément central de la mise au point, du déploiement, de l'utilisation et du contrôle des systèmes d'IA, en veillant au respect des droits fondamentaux, y compris ceux consacrés par les traités de l'Union européenne et la charte des droits fondamentaux de l'Union européenne, qui se rejoignent tous dans un fondement commun ancré dans le respect de la dignité humaine, en vertu duquel l'être humain jouit d'un statut moral unique et inaliénable. Cela implique également la prise en compte de l'environnement naturel et des autres êtres vivants qui font partie de l'écosystème humain, ainsi qu'une approche durable permettant l'épanouissement des générations à venir »⁹³.

L'approche européenne cherche à demeurer cohérente avec la volonté de bénéficier de standards de protection contre les risques sociaux de haut niveau. Il s'agit notamment de la protection de la vie privée, des données et des discriminations.

2.2.2. Discussion autour des SALA

Jusque très récemment, l'Union européenne a plus ou moins ignoré, dans les travaux sur l'Intelligence artificielle, les éléments relatifs à la défense et la sécurité. Ce n'est qu'au cours de l'année 2019 qu'elle a commencé – timidement – à les traiter⁹⁴. Ce changement est notam-

⁹² « Ethics washing made in Europe », *Der Tagesspiegel*, 08 avril 2018.

⁹³ Groupe d'experts indépendants de haut niveau sur l'Intelligence artificielle constitué par la Commission européenne en juin 2018, *Lignes directrices en matière d'éthique pour une IA digne de confiance*, avril 2019, p. 49.

⁹⁴ Franke (U. E.), *Not smart enough: The poverty of European military thinking on artificial intelligence*, European Council on Foreign Relations, Policy Brief, December 2019.

ment survenu parce que la Finlande a décidé d'inclure la thématique à l'agenda des institutions européennes durant sa Présidence du Conseil. La problématique des SALA est cependant une exception puisqu'elle a fait l'objet d'un traitement abondant et ancien (2013). En septembre 2018, deux textes ont notamment été votés par le Parlement européen sur la thématique des systèmes d'armes autonomes.

Le premier texte demandait aux États-membres et au Conseil européen de travailler à la création d'une interdiction internationale concernant les systèmes d'armes dont le fonctionnement manque d'un contrôle humain significatif sur la fonction critique qu'est la sélection et l'engagement des cibles⁹⁵. Il s'agissait d'une motion, c'est-à-dire d'une simple demande pour que celle-ci puisse faire l'objet d'un processus formel, sous la forme d'une résolution, au sein du Parlement. Les motions ne sont pas des textes juridiquement contraignants.

Le second texte correspond en quelque sorte à la formalisation demandée. Il s'agit d'une résolution. Le Parlement européen publie régulièrement des résolutions. Il s'agit de textes qui n'ont pas non plus de caractère contraignant, mais qui disposent d'une forte valeur déclarative et symbolique. Ils permettent au Parlement, sur des questions internationales, de présenter son avis sur des sujets d'actualité. Cette résolution indique que les parlementaires considèrent qu'un contrôle humain significatif sur les fonctions critiques des systèmes d'armes doit toujours être préservé⁹⁶. Elle ne reprend donc pas la demande de participation des États-membres et du Conseil européen à la naissance d'un dispositif d'interdiction internationale.

Les ministres en charge de la Défense des 28 États-membres ont décidé de mettre à l'agenda de leurs réunions la thématique des SALA. Le sujet a fait l'objet d'une discussion informelle en janvier 2019 à Bucarest. Il a ensuite été officiellement abordé lors de leur rencontre d'août à Helsinki.

À cette occasion, la Commission a mis en œuvre une initiative un peu inhabituelle. La veille de la réunion des ministres, pour préparer leurs discussions, elle a organisé un dîner informel auquel ont participé des membres du *Global Tech Panel*. Ces experts ont expliqué en quoi l'Intelligence artificielle et la digitalisation vont impacter le monde de la défense et de la sécurité, puis ont échangé avec les ministres. La problématique des SALA a bien évidemment été abordée. En organisant cette rencontre informelle, la Commission a cherché à démontrer que le traitement des défis que pose le développement de l'Intelligence artificielle ne peut se faire que grâce à un partenariat entre les autorités gouvernementales, les acteurs des nouvelles technologies et la société civile⁹⁷.

⁹⁵ European Parliament, *Motion for a Resolution on Autonomous Weapon Systems*, September 5, 2018.

⁹⁶ European Parliament, *Resolution of 12 September on Autonomous Weapon Systems, 2018/2752(RSP)*, September 12 2018.

⁹⁷ « EU defence ministers discuss AI with Global Tech Panel experts », *European External Action Service*, 4 September 2019.

2.3. Organisation de coopération et de développement économiques (OCDE)

En mai 2019, l'OCDE a publié ses principes pour l'IA. Il s'agit des premiers standards internationaux ayant obtenu une reconnaissance par les gouvernements⁹⁸. Ils incluent des recommandations politiques pratiques, mais également des principes fondés sur des valeurs, pour mettre en œuvre une gestion responsable d'une IA digne de confiance. Le texte pose ainsi que :

- ➔ L'Intelligence artificielle doit bénéficier aux populations et à la planète en permettant de stimuler la croissance inclusive, le développement durable et le bien-être ;
- ➔ Les systèmes recourant à l'Intelligence artificielle doivent respecter les règles de droit, en particulier les droits de l'Homme, les valeurs démocratiques et la diversité. Ils doivent disposer des dispositifs de protection appropriés afin d'être équitables ;
- ➔ Une transparence doit exister à propos de l'Intelligence artificielle afin d'assurer que les personnes comprennent les conséquences de l'emploi des systèmes dotés d'IA et puissent éventuellement les contester ;
- ➔ Les systèmes d'IA doivent fonctionner de manière robuste, sûre et sécurisée tout au long de leurs cycles de vie. Les risques les concernant doivent être évalués en permanence ;
- ➔ Les organisations et les individus concevant, déployant et opérant des systèmes dotés d'Intelligence artificielle doivent pouvoir être tenus responsables.

En 2019, ces principes avaient été approuvés par les gouvernements des 36 États-membres de l'Organisation, ainsi que par l'Argentine, le Brésil, la Colombie, le Costa-Rica, le Pérou et la Roumanie. Les principes de l'OCDE ont par ailleurs été largement repris par le G20 pour concevoir ces propres principes centrés sur l'humain.

Afin notamment d'étudier les méthodes permettant de mettre en œuvre ces principes, l'OCDE a créé un *AI Policy Observatory*. Constitué au moment où les Principes étaient adoptés, il sert à mettre à disposition des ressources venant de l'Organisation, mais également de ses partenaires et de toutes les parties prenantes. Cette plateforme facilite ainsi le dialogue entre tous les intervenants aux débats concernant les Principes et leur application.

L'Observatoire travaille avec les communautés politico-administratives, mais également avec les acteurs de l'économie digitale, des sciences et technologies, de la santé, de l'emploi, de la protection des consommateurs, de l'éducation, des transports... afin de pouvoir analyser très largement les opportunités liées aux développements futurs de l'Intelligence artificielle, ainsi que les défis qu'ils représentent. Il dispose d'un centre au sein duquel des données factuelles (séries statistiques, rapports et publications de travaux de recherche...), des « preuves », sur l'IA sont collationnées et partagées. Cette bibliothèque comprend également des études et rapports sur les politiques nationales et les potentiels développements sectoriels de l'IA (santé, transport...). Il sert également de forum entre les parties prenantes (représentants des

⁹⁸ Ces principes ont en effet été adoptés par les États-membres lorsque ceux-ci ont approuvé une Recommandation du Conseil sur l'intelligence artificielle.

gouvernements, des communautés académiques, du secteur privé, de la société civile, d'autres organisations internationales...).

2.4. G7

Lors d'une réunion du G7 au Canada en juin 2018, la présidence canadienne a présenté aux autres membres (France, Allemagne, Italie, Japon, Royaume-Uni et États-Unis d'Amérique) ses travaux sur une Vision commune pour le futur de l'intelligence artificielle. Ce document, discuté par les ministres en charge de l'innovation des 7 pays, a été adopté. La Vision commune comprend 12 engagements que les membres du G7 doivent respecter :

- S'efforcer de promouvoir une Intelligence artificielle centrée sur l'humain et son adoption commerciale, tout en continuant d'avancer pour que soient développées des approches neutres d'un point de vue technique, éthique et technologique⁹⁹ ;
- Promouvoir les investissements pour la R&D permettant de générer dans le public une confiance des nouvelles technologies et encourager les industries à investir dans le développement et le déploiement d'une Intelligence artificielle soutenant une croissance durable et une autonomisation économique des femmes (en traitant des thématiques de l'imputabilité, des assurances, de la mise en responsabilité, de la sécurité, de la sûreté, des biais et potentiels mésusages) ;
- Soutenir la formation tout au long de la vie, l'éducation, l'entraînement et l'acquisition de nouvelles qualifications et l'échange pour le développement au sein de la force de travail des compétences en IA (notamment au travers de l'apprentissage) ;
- Soutenir et impliquer les femmes, les groupes sociaux sous-représentés et les individus marginalisés afin qu'ils participent à tous les stages du développement et de la mise en œuvre de l'Intelligence artificielle (en tant que créateurs, parties prenantes, autorités prenant les décisions) ;
- Faciliter un dialogue entre les parties prenantes sur la manière de générer de l'innovation en matière d'IA tout en accroissant la confiance et l'adoption, et fournir des informations pour les discussions politiques futures ;
- Soutenir les efforts pour promouvoir la confiance dans l'IA, avec une attention particulière à la lutte contre les stéréotypes néfastes et au renforcement de l'égalité entre les genres. Soutenir les initiatives qui promeuvent la sûreté et la transparence et fournir des lignes directrices concernant l'intervention humaine dans les procédures de prise de décision des IA ;
- Promouvoir les usages de l'IA au sein des entreprises, en particulier les PME et les firmes qui ne font pas partie de secteurs *hi-tech* ;

⁹⁹ Par des dispositifs relativement bien précisés par le texte : sauvegardes de la vie privée (développement de régimes légaux appropriés, investissements dans la cybersécurité, mise en vigueur appropriée des législations existantes sur la protection de la vie privée et communication officielle autour de cette mise en vigueur) ; information des individus sur les corpus juridiques nationaux (notamment ceux sur la protection des données à caractère personnel) ; promotion de la R&D industrielle sur la sûreté, les assurances, la qualité des données et la sécurité des données ; exploration des emplois d'autres technologies transformatives.

- Promouvoir les politiques volontaires pour le marché du travail, le développement de la force de travail et les programmes de reconversion professionnelle, afin de disposer des compétences nécessaires pour les nouveaux métiers et pour éviter que des travailleurs ne soient mis de côté ;
- Encourager les investissements en IA et pour l'innovation afin de créer de nouvelles opportunités pour l'ensemble des populations ;
- Encourager les initiatives, notamment celles des entreprises, permettant de renforcer la sécurité digitale, de développer des codes de conduite, standards et lignes directrices et de partager les meilleures pratiques ;
- Assurer que le design de l'IA et sa mise en œuvre respectent les cadres applicables à la protection de la vie privée et des données, et les promeuvent ;
- Soutenir un environnement commercial ouvert et équitable pour que l'accès aux flux de données soit libre, tout en respectant les cadres de protection de la vie privée et des données (en traitant les questions des pratiques commerciales discriminatoires, de la localisation injustifiée des données et en reconnaissant la protection des droits de propriété intellectuelle).

À cette occasion, le Canada, qui présidait à cette époque le G7, a proposé à ses partenaires d'organiser un Sommet international sur l'intelligence artificielle réunissant États, entreprises et chercheurs.

2.5. Global Partnership on Artificial Intelligence (GPAI)

Parallèlement à l'adoption de la Vision commune (voir ci-avant), le Canada et la France ont décidé d'annoncer le lancement d'une initiative bilatérale sur l'IA lors de la réunion du G7 de juin 2018. Dans ce cadre, les deux pays ont décidé de la création d'un Groupe international d'experts sur l'intelligence artificielle (G2IA).

Le mandat du G2IA a été présenté, en décembre 2018, par les deux pays lors de la conférence multipartite du G7 sur l'intelligence artificielle. Le Groupe international a pour mission d'aider la communauté internationale à concevoir des cadres permettant de développer l'IA tout en préservant les droits de l'homme, l'inclusion, la diversité, l'innovation et la croissance économique. De la sorte, il doit favoriser la coopération et la coordination des différentes politiques nationales de promotion de l'Intelligence artificielle.

En mai 2019, les autorités des deux pays ont précisé l'organisation interne du Groupe et ses missions lors d'une réunion informelle des ministres du G7 en charge du numérique. Au cours de cette réunion, l'Allemagne, l'Italie, l'Inde, le Japon, le Royaume-Uni, la Nouvelle-Zélande et l'Union européenne ont manifesté leur intérêt concernant la structure.

Les travaux du Groupe ont finalement été officiellement lancés en août lors du Sommet du G7 de Biarritz d'août 2019.

La Canada et la France avaient également annoncé, en décembre 2018, le lancement d'un Partenariat mondial sur l'intelligence artificielle (PMIA). Lors du Sommet du G7 à Biarritz en août 2019, les autres membres du G7 ont pris note de l'existence de cette initiative.

En mai 2020, lors de la réunion des ministres Science & Technologie du G7, les pays du G7 se sont accordés pour lancer le PMIA. L'objectif est de renforcer la coopération multi-acteurs pour que la progression de l'intelligence artificielle intègre leurs valeurs démocratiques et réponde aux défis mondiaux. Dans un premier temps, il a notamment été décidé de concentrer l'attention du Partenariat sur les réponses à apporter à la pandémie de Covid-19. Dans la lignée des textes adoptés conjointement par les membres du G7 auparavant (notamment la Vision commune), ils se sont engagés envers un développement et une utilisation responsables d'une IA centrée sur l'humain, respectant les droits de l'homme, les libertés fondamentales et les valeurs démocratiques communes.

Ils ont été rejoints par l'Australie, l'Inde, le Mexique, la Nouvelle Zélande, la République de Corée, Singapour, la Slovénie et l'Union européenne. L'ensemble de ces acteurs ont rendu publique, en juin 2020, une déclaration commune permettant de lancer véritablement le Partenariat. Une autre a porté sur le financement de l'initiative. Celle-ci vise à soutenir et guider le développement responsable de l'IA – ce qui inclut un développement respectueux de la protection des droits de l'Homme, l'inclusion, la diversité, l'innovation et la croissance économique. Elle sert à rassembler des parties prenantes diverses et cherche à combler le fossé entre théorie et pratique en soutenant les recherches de haut niveau et les activités applicatives sur certaines priorités.

Le GPAI est fondé sur une mise en relation d'experts reconnus issus de l'industrie, de la société civile, des administrations nationales et des structures académiques. Certains d'entre eux sont rassemblés au sein du G2IA. Ils doivent collaborer dans des groupes de travail thématiques. Quatre thèmes ont été retenus : 1/ Intelligence artificielle responsable ; 2/ Gouvernance des données ; 3/ Futur du travail ; 4/ Innovation et commercialisation.

Le GPAI est soutenu par l'OCDE. Il dispose d'un secrétariat à Paris et de 2 centres d'expertise à Montréal et Paris. Les rapports avec l'OCDE sont relativement étroits. Les 2 centres fournissent un soutien administratif et scientifique aux projets pratiques pris en charge ou supervisés par des groupes de travail pluridisciplinaires constitués d'experts.

3. Termes des débats

3.1. Thématiques faisant l'objet d'un traitement

Dans leur analyse comparée de 84 documents, nationaux et internationaux, contenant des principes ou des lignes directrices éthiques concernant l'Intelligence artificielle, Jobin, Ienca et Vayena ont répertorié les principaux concepts traités¹⁰⁰.

¹⁰⁰ Jobin (A.), Ienca (M.) et Vayena (E.), « The global landscape of AI ethics guidelines », *op. cit.*

Ethical principle	Number of documents	Included codes
Transparency	73/84	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing
Justice & fairness	68/84	Justice, fairness, consistency, inclusion, equality, equity, (non-)bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-maleficence	60/84	Non-maleficence, security, safety, harm, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility	60/84	Responsibility, accountability, liability, acting with integrity
Privacy	47/84	Privacy, personal or private information
Beneficence	41/84	Benefits, beneficence, well-being, peace, social good, common good
Freedom & autonomy	34/84	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment
Trust	28/84	Trust
Sustainability	14/84	Sustainability, environment (nature), energy, resources (energy)
Dignity	13/84	Dignity
Solidarity	6/84	Solidarity, social security, cohesion

Ces auteurs précisait qu'il n'existait pas – en 2019 – de thématiques apparaissant dans tous les documents. Ils constataient cependant qu'une convergence existe à propos de la transparence, de la justice et de l'équité, de la non-malfaisance et de la responsabilité. Si ces thèmes sont traités dans la plupart des publications, ils ne font toutefois pas l'objet d'une unité : des différences sémantiques et conceptuelles sont présentes, qui expliquent que les recommandations ne soient parfois pas les mêmes.

► Transparence

Bien qu'il s'agisse de la thématique la plus souvent traitée dans les documents sur les dimensions éthiques de l'IA, la transparence fait l'objet d'approches très diversifiées. L'interprétation donnée à la notion, les justifications concernant sa protection, son domaine d'application et les modalités de sa mise en œuvre sont traités avec des différences parfois significatives.

Les références à la transparence comprennent les actions visant à accroître l'explicabilité ou l'interprétabilité, mais également les actes de communication et de divulgation d'informations concernant les technologies d'IA utilisés par les systèmes. Les domaines principaux auxquels est rattachée la notion sont l'utilisation des données (en particulier personnelles), les interactions homme-machines dotées d'IA, les décisions automatisées et la finalité de l'utilisation des données et du système d'IA. Assez souvent, la transparence est conçue comme

l'une des voies permettant de renforcer la confiance. Les traductions juridiques de la transparence constituent également une thématique particulièrement investiguée (en particulier la création de nouveaux droits individuels).

► **Justice, impartialité et équité**

Dans les documents concernant l'éthique de l'IA, la notion de justice est souvent ramenée à l'idée d'impartialité, à la prévention, au suivi et à l'atténuation des biais indésirables ainsi qu'aux problématiques de discrimination. Certains textes proposent des traductions juridiques du concept (voir ci-après).

Des références ont notamment souligné l'importance d'un accès équitable à l'Intelligence artificielle, aux données qui permettent de développer des systèmes et/ou aux bénéfices engendrés grâce à l'IA.

De même, la transparence est souvent présentée comme une réponse aux difficultés pour disposer d'une IA juste, impartiale et équitable. Il s'agit notamment de fournir des informations compréhensibles et de sensibiliser le public sur les risques, mais également les réglementations et les droits existants.

Les analyses produites par les institutions publiques replacent souvent la question de la justice dans un cadre socioéconomique, en particulier la mette en lien avec le marché du travail (capacité différenciée des travailleurs à comprendre et employer les systèmes dotés d'IA).

Certaines sources se concentrent sur les risques liés aux biais contenus dans les jeux de données employés pour entraîner, puis faire fonctionner les outils d'IA. Elles soulignent la nécessité de disposer de données précises, complètes et diversifiées.

► **Non-malfaisance**

Dans les références produites, au niveau national comme à l'international, le thème de la non-malfaisance intègre souvent les questions de la sécurité et de la sûreté, ainsi que, plus globalement, l'idée que les systèmes d'IA doivent être conçus pour disposer d'un état dans lequel ils ne peuvent provoquer aucun préjudice, prévisible comme non intentionnel. Plus précisément, les réflexions portent souvent sur la matérialisation de risques spécifiques ou la survenue de préjudices particuliers (comme la mauvaise utilisation intentionnelle des systèmes dotés d'IA grâce à des cyberattaques) et les stratégies de gestion du risque qui pourraient être mises en œuvre. Les dommages souvent pris en considération dans les références sont la discrimination, la violation de la vie privée ainsi que les dommages corporels. D'autres nuisances sont parfois abordées : perte de compétences professionnelles, insuffisance des mesures réglementaires, impacts négatifs de long terme sur le bien-être social, conséquences psychologiques et émotionnelles délétères pour les individus...

L'un des sujets désormais traités au sein de certaines des enceintes internationales est en effet l'impact éventuel de l'Intelligence artificielle sur les expériences émotionnelles humaines. Il s'agit notamment de la capacité des dispositifs utilisant de l'IA à intégrer les sensibilités culturelles des individus et des communautés. Les souffrances émotionnelles sont perçues comme un risque, en particulier pour les dispositifs – comme les robots sociaux – qui vont entretenir des relations intimes avec les humains.

Quelques documents citent explicitement, parmi les risques potentiels, les détournements liés au double usage.

► Responsabilité

La responsabilité est une thématique que de nombreuses initiatives, nationales et internationales, entendent traiter. Toutefois, en dépit de fréquentes références à une « IA responsable », la responsabilité et la mise en responsabilité sont rarement précisément définies.

Dans ce domaine, des textes font de l'intégrité l'une des composantes premières de la responsabilité.

Certaines recommandations demandent aussi que soit systématiquement opérée une clarification des responsabilités des différents acteurs concernés par la conception et la mise en œuvre d'un système doté d'Intelligence artificielle (en particulier les responsabilités juridiques), si possible dès l'origine. Les acteurs cités comme pouvant être responsables et comptables pour les actions et décisions d'une IA sont les développeurs de l'Intelligence artificielle, les *designers* du système qui l'emploie, les organisations utilisatrices et les industries.

Beaucoup de références se concentrent sur le besoin de rendre l'Intelligence artificielle auditable. Il s'agit de disposer des moyens pour s'assurer que la responsabilité des producteurs, concepteurs et propriétaires/utilisateurs des systèmes utilisant de l'IA puisse être mise en cause en cas de préjudices.

Une autre approche proposée consiste à se concentrer sur les raisons et processus qui pourraient amener des impacts négatifs, et donc une mise en cause de la responsabilité. Cette manière de traiter la question constitue un complément à la non-malfaisance.

Bien évidemment, la question de la responsabilité doit être mise en relation avec celle de l'autonomie conférée à la machine. Lors des discussions qui ont permis de préparer les lignes directrices et standards éthiques, des oppositions sont parfois apparues concernant le fait qu'une Intelligence artificielle puisse être rendue seule responsable comme un humain pourrait l'être. Certains acteurs ont ainsi affirmé que des humains devraient toujours être les seuls acteurs pouvant finalement être mis en responsabilité pour l'emploi des artefacts technologiques.

Plus globalement, la plupart des cycles de réflexion lancés au niveau international ont permis de reconnaître que l'engagement du public dans les débats devait être accentué. De même, beaucoup concluent qu'au regard des risques potentiels liés à l'IA, un effort doit être réalisé en matière d'éducation.

3.2. Considérations juridiques

La plupart des textes produits sur l'IA éthique ont abordé la question de la traduction des principes en règles de droit. Certains documents se concentrent par exemple sur une vision très juridique du concept de justice. Celle-ci est alors définie comme l'application des normes existantes en matière de diversité, une obligation de l'inclusion et le respect de l'égalité. Quelques textes considèrent par ailleurs qu'un droit à faire appel des décisions sous-tendues

par des systèmes d'IA, voire à les contester, devrait être créé. De même, certains jeux de principes éthiques proposent que des droits à réparation et à correction soient instaurés.

En matière de transparence, le raisonnement est parfois similaire. Pour que celle-ci puisse être assurée, certains textes demandent qu'un droit à l'information soit créé concernant les fonctionnalités, les méthodes de développement, les données utilisées et le type d'IA des systèmes. Certains documents prônent même la création d'une obligation juridique à ce que les personnes soient informées dès lors qu'un dispositif recourant à de l'Intelligence artificielle va leur être appliqué.

Dans un certain nombre d'exercices de réflexion, les questions juridiques ont intégré le statut des systèmes d'IA. Des acteurs comme l'*Institute of Electrical and Electronics Engineers* (IEEE) se sont clairement exprimés contre l'attribution d'une personnalité juridique individuelle aux machines.

3.3. Réflexions sur les outils à mettre en place

Lorsque les documents proposent des solutions pour s'assurer que la justice et l'équité puissent être respectées, il ne s'agit pas toujours d'outils juridiques (voir ci-avant). Beaucoup de sources mettent en avant des solutions techniques, en particulier les standards, ou du codage normatif explicite.

De même, les lignes directrices concernant la prévention des dommages présentes dans les documents se concentrent sur la création d'ensembles de mesures techniques et la mise en œuvre de stratégies de gouvernance. Elles comprennent des interventions aux niveaux de la recherche en IA, du *design*, du déploiement... (solutions techniques permettant l'évaluation de la qualité des données, renforçant la sécurité des données, *privacy by design*...). Les stratégies en matière de gouvernance qui sont proposées incluent la mise en place de dispositifs permettant la coopération active entre les spécialistes de disciplines différentes et entre les parties prenantes, la conformité par rapport aux normes existantes, ainsi que la création de procédures et pratiques de supervision (en particulier les tests, la surveillance, les audits et les évaluations par des entités internes, les consommateurs, les utilisateurs, des tierces parties indépendantes ou des organismes publics). Les approches ainsi proposées reposent généralement sur la production de standards pour la mise en œuvre de l'IA et l'évaluation des résultats qu'elle produit.

La plupart des initiatives internationales lancées ont donc permis d'aboutir au constat que de nouveaux standards sont nécessaires, qui préciseront les niveaux mesurables et testables de transparence des systèmes pour lesquels la conformité devra pouvoir être objectivement évaluée. Beaucoup d'intervenants considèrent que les chercheurs et les concepteurs doivent par ailleurs opérer avec une culture de la sécurité permanente, de manière à préempter les comportements inattendus ou non-anticipés des systèmes sur lesquels ils travaillent.

Diverses sources demandent par ailleurs que les autorités développent des procédés permanents de surveillance. Dans certains documents, cette surveillance est conçue comme devant être interactive, pluridisciplinaire et inclure de nombreuses parties prenantes, y compris des acteurs issus de la société civile.

4. Travaux au sein de l'OTAN

Bien que la thématique de l'Intelligence artificielle n'ait jamais été mise à l'agenda d'un Sommet de l'OTAN, les responsables politiques et militaires de l'Organisation y ont été confrontés car le Conseil de l'Atlantique nord et le Comité militaire ont décidé que leur *Away Day* 2018 serait consacré aux technologies disruptives, en particulier à l'Intelligence artificielle.

Par ailleurs, différentes entités au sein de l'OTAN ont lancé, ces dernières années, des activités en rapport avec l'IA ou ont inclus la thématique à leurs activités courantes :

- ▶ **L'Organisation de l'OTAN pour la science et la technologie (STO) a fait des applications de l'IA et du Big data pour la prise de décision, ainsi que de l'autonomie, des thèmes prioritaires en matière de science et technologie¹⁰¹.**

En janvier 2017, l'OTAN a adopté le concept d'« approche thématique ». Il s'agit de traiter de sujets pluridisciplinaires qui influent sur les capacités militaires et que les armées des États-membres de l'Organisation doivent pouvoir maîtriser. L'adoption d'un thème permet bien évidemment de stimuler les activités de la STO le concernant (notamment d'identifier les manques en matière de recherche scientifique et technologique). Elle a aussi pour effet de faciliter la collaboration entre les panels de la STO pour qu'ils traitent ce sujet.

Dès 2017, le thème de « l'IA et du *Big data* pour la prise de décision » a commencé à être traité¹⁰². Il a été conçu très largement puisqu'il intègre notamment les outils et méthodes d'analyse de l'information, les architectures, l'entraînement des systèmes et la visualisation des résultats qu'ils produisent, la guerre informationnelle et le support à la décision. Une réunion de spécialistes s'est notamment tenue à Bordeaux en juin 2018.

L'« autonomie à partir d'une perspective système » est un autre thème qui a été choisi dès 2017. Une réunion de spécialistes, organisée en Norvège en mai 2017, a permis d'identifier 17 domaines d'intérêt en lien avec cette thématique (domaines pouvant faire l'objet de travaux). Cette rencontre a ouvert une séquence préparatoire assez longue au cours de laquelle la pertinence de ces domaines a été interrogée, leurs périmètres précisés et les types d'actions pour les traiter définis. Ce travail se poursuit.

¹⁰¹ La STO est chargée, au sein de l'OTAN, de produire de l'innovation. Elle est également une structure de conseil et d'expertise scientifique. Il s'agit avant tout d'un forum collaboratif qui aide l'Alliance au travers d'actions de recherche scientifique et technologique (300 projets soutenus par an en moyenne) et qui rassemble une communauté de plus de 6 000 scientifiques.

Le Comité OTAN pour la science et la technologie (STB) a pour mission de superviser directement les activités de la STO et d'assurer une gouvernance unifiée du secteur S&T de l'OTAN. Il est ainsi chargé d'assurer la cohérence de l'ensemble de ce secteur. Il a pour fonction de définir des objectifs dans le cadre de la stratégie S&T de l'OTAN, notamment de fixer les priorités (afin d'orienter les travaux financés ou cofinancés par la STO). Il joue aussi un rôle de point de coordination pour l'ensemble des programmes de travail.

¹⁰² Smallegange (A.), Venema (A.), Bastiaansen (H.), Bronkhorst (A.), « Big Data and Artificial Intelligence for Decision Making: Dutch Position Paper », STO-MP-IST-160, 2018.

Le fait que « l'IA et le *Big data* pour la prise de décision » et l'« autonomie à partir d'une perspective système » aient été choisies comme approches thématiques a ainsi amené l'organisation de rencontres entre spécialistes. Ce choix a aussi eu pour conséquence de multiplier les activités intégrant de l'IA dans le Programme de travail collaboratif de la STO¹⁰³.

Dans ce cadre, différentes actions concernant les dimensions éthiques (ou des thématiques liées) ont été réalisées ou planifiées. Un atelier de recherche va par exemple être organisé sur le thème du « *Meaningful Human Control of AI-based Systems (Key Characteristics, Influencing Factors and Design Considerations)* ». De même, un *Research Task Group* (RTG – une équipe de chercheurs menant des analyses techniques) a été constitué pour traiter de la question de l'« *Human Systems Integration for Meaningful Human Control over AI-based systems* ». Sur-tout, un autre RTG doit s'attaquer au problème des « *Ethical, legal and moral (ELM) impacts of novel technologies on NATO's operational advantage* », qui intégrera des analyses à propos de l'IA.

- ▶ **Le Commandement allié pour la transformation (*Allied Command Transformation – ACT*) organise différents événements consacrés à l'Intelligence artificielle.**

Ces événements ont notamment porté sur les opportunités et les défis que les technologies peuvent représenter. Durant les Forums NATO-Industrie, ainsi que les Conférences internationales sur le développement et l'expérimentation des concepts des dernières années, l'Intelligence artificielle et des thèmes qui lui sont liés ont par exemple suscité des présentations et des discussions (y compris sur les dimensions éthiques).

- ▶ **La *NATO Communications and Information Agency* (NCI Agency) a intégré l'Intelligence artificielle aux programmes de certains des événements qu'elle organise.**

L'Agence a notamment fait de l'Intelligence artificielle un sujet central du *NATO Information Assurance Symposium* de 2018. De même, elle a organisé un défi dans le cadre du *Hackathon for Good* de novembre 2018¹⁰⁴. L'Intelligence artificielle a par ailleurs été l'un des thèmes de la NITEC19 organisée à Oslo en mai 2019¹⁰⁵.

¹⁰³ NATO Science and Technology Organization, *Collaborative Program of Work and budget for year 2020*. Voir annexes 1 et 2.

¹⁰⁴ Le Hackathon pour la Paix, la Justice et la Sécurité (« *Hackathon for Good* ») est organisé chaque année depuis 2018. Accueilli à La Haye par la *Hague Data Science Initiative*, il permet au cours d'un week-end aux participants de travailler sur des thèmes décidés par la *Hague Data Science Initiative*, certaines ONG et des organisations internationales. Les équipes doivent proposer des solutions pour répondre aux problèmes posés par les organisateurs. En 2018, l'un des 6 thèmes retenus a été celui de la *NCI Agency* : l'utilisation des technologies pour identifier les actions de propagande et les informations « inflammatoires ». L'Agence a demandé aux 5 équipes participantes de développer un outil capable de vérifier des objets et des textes dans des images et de déterminer s'il s'agissait d'éléments de propagande. Pour ce faire, elle leur a fourni un jeu d'images d'entraînement.

¹⁰⁵ La NITEC est la conférence annuelle organisée par la *NCI Agency* qui permet de rassembler les industriels des pays de l'OTAN travaillant dans le secteur de la Défense. En 2019, elle a traité plus spécifiquement des solutions technologiques avancées et des pratiques commerciales spécifiques qu'elles nécessitent.

► **Le NATO Industrial Advisory Group (NIAG) s'est également engagé dans le traitement des problématiques liées à l'IA¹⁰⁶.**

Entre autres travaux, il a ainsi publié deux études relatives à l'utilisation du *Big data* par l'OTAN et à l'impact de l'autonomie sur les processus de planification et les opérations de l'Alliance en 2017 et 2018¹⁰⁷.

Les actions de promotion de l'IA de l'Alliance ont aussi comporté des expérimentations. En octobre et novembre 2018, l'exercice *Trident Juncture* a ainsi intégré une vingtaine d'expérimentations en condition quasi-opérationnelle¹⁰⁸.

Les efforts internes ont été accompagnés de l'organisation d'initiatives communes au niveau international. En novembre 2019, la Présidence finlandaise du Conseil de l'UE et l'*EU Institute for Security Studies* ont co-organisé à Bruxelles une conférence sur les relations Union européenne-OTAN et l'Intelligence artificielle¹⁰⁹. L'événement a notamment permis de réfléchir aux potentielles actions de coopération qui pourraient être lancées. Au cours de la conférence, une partie des discussions a porté sur les efforts de régulation de l'IA et de lutte contre sa dissémination.

Enfin, lors d'une discussion de haut niveau en ligne organisée récemment (octobre 2020)¹¹⁰, le Secrétaire général adjoint de l'OTAN Mircea Geoană a affirmé que l'Alliance devait pouvoir jouer un rôle de facilitateur pour que la coopération transatlantique, notamment entre l'OTAN et l'UE, s'affermisse concernant l'Intelligence artificielle. Elle pourrait soutenir l'innovation, tout en facilitant les échanges.

5. Efforts de standardisation au niveau international

5.1. Prise en charge par l'International Organization for Standardization (ISO)

Au sein de l'*International Organization for Standardization*, un dispositif dédié à l'Intelligence artificielle a été mis en place. Le comité en charge de la standardisation dans le domaine des technologies de l'information (ISO/IEC JTC 1) s'est ainsi doté d'un sous-comité (SC 42) en 2017.

¹⁰⁶ Créée en 1968, le NIAG est un organisme consultatif d'expertise technique de haut niveau. Composé de représentants des industries des pays membres et des pays partenaires de l'OTAN, il conseille l'Alliance, et notamment la Conférence des directeurs nationaux des armements, sur des questions clés relatives à la coopération en matière d'armement et à la base industrielle et technologique de l'Alliance.

¹⁰⁷ NATO Industrial Advisory Group Study Group 208, final report « On adopting big data in NATO », AC/322-N(2017)0085, June 2017 et Study Group 231 "Autonomy's impact on future NATO Planning and Operations – Perspectives from the Industry".

¹⁰⁸ Tucker (P.), « How NATO's Transformation Chief Is Pushing the Alliance to Keep Up in AI », DefenseOne, May 18, 2018.

¹⁰⁹ Pour un compte-rendu de la conférence, voir <https://www.iss.europa.eu/sites/default/files/EUISSFiles/EU%20NATO%20AI%20-%20Report.pdf>.

¹¹⁰ Cette discussion portait sur la coopération transatlantique dans le domaine de l'IA. Organisée par l'*Atlantic Council* et le *GeoTech Center*, elle a notamment permis au Secrétaire général adjoint de discuter avec le président et le vice-président de la *National Security Commission on Artificial Intelligence* américaine, ainsi qu'avec l'ancien Secrétaire à la Défense adjoint Robert O. Work.

Il est notamment chargé de fournir des cadres et des orientations aux comités de l'ISO et de l'*International Electrotechnical Commission* (IEC) qui produisent des normes pour les applications fondées sur de l'IA¹¹¹.

Ce sous-comité est, au sein des deux organisations, la seule structure travaillant pour l'ensemble de l'écosystème de l'IA. C'est pourquoi il doit notamment intégrer les contributions de parties prenantes nombreuses et diversifiées. Son président, Wael William Diab, avait par ailleurs précisé qu'il existait un fossé entre ce qu'est l'IA actuellement et ce qu'elle paraît être pour certains. Les travaux de la structure doivent notamment permettre de rendre compte de l'ensemble des nombreuses technologies qui constituent l'Intelligence artificielle. Ils doivent aussi servir à éviter que l'IA ne soit trop facilement assimilée à l'autonomie complète des systèmes dans lesquels elle est implantée. Dès le début des travaux du sous-comité, son président a indiqué que les utilisateurs devaient pouvoir comprendre comment les dispositifs d'IA prennent leurs décisions. Il considérait donc que des aspects comme la transparence et l'explicitabilité des systèmes sont fondamentaux, nécessitant une standardisation globale.

Les travaux du sous-comité ont notamment débuté par la production de standards de base, en particulier des éléments de terminologie, de classification et de définition des concepts. L'objectif était de faire bénéficier toutes les parties prenantes d'un vocabulaire commun.

Un autre travail préliminaire a consisté à produire une évaluation des différentes approches en matière de traitement et de préciser les caractéristiques des principales catégories de systèmes d'IA. Il impliquait une étude des différentes technologies (algorithmes de *Machine learning*, méthodes de raisonnement...) utilisées par les systèmes¹¹². Il s'agissait aussi d'analyser les systèmes d'IA spécialisés existants, afin d'identifier et comprendre leurs caractéristiques, ainsi que les approches en termes de calcul et les architectures qui sous-tendent leur fonctionnement.

Enfin, le dernier travail préparatoire concernait la loyauté/confiance. Tous les éléments permettant de les obtenir ont été étudiés (sécurité, respect de la vie privée, robustesse, transparence, limitation des biais). Il a ainsi été possible de produire des séries de recommandations pour les projets de standardisation.

Plus tardivement, le sous-comité a commencé à travailler sur les « enjeux de société ». Les technologies comme l'Internet des objets et l'IA ont la capacité d'influencer la manière de vivre des générations à venir. Il s'agit donc pour le sous-comité de prendre en compte les considérations socio-économiques (comme les transformations de la force de travail) qui risquent de survenir et surtout les éléments qui peuvent les influencer (biais des algorithmes, conditions de sécurité en milieu industriel...). Pour traiter ces dimensions éthiques, le SC 42 travaille avec d'autres comités au sein des deux organisations mères.

L'organe comprend 31 membres et 16 observateurs. Différents groupes internes de travail ont ainsi été constitués.

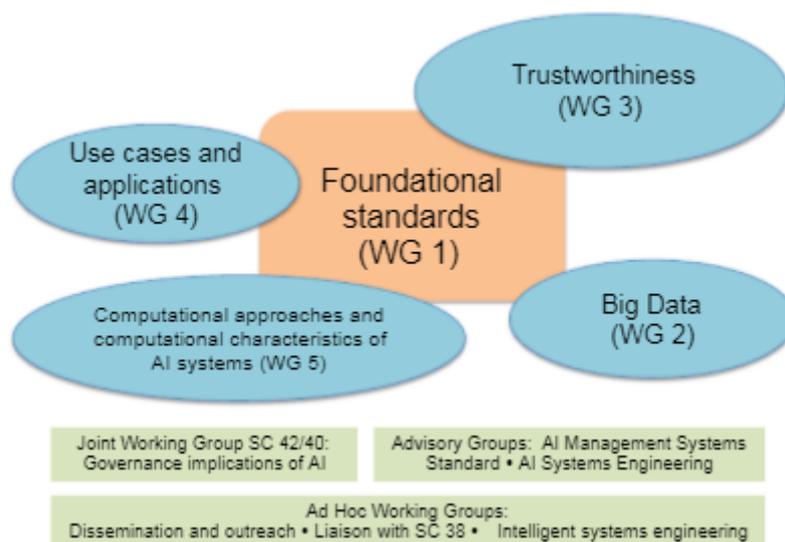
¹¹¹ Le SC 42 est rattaché à la fois à l'ISO et à l'IEC.

L'ISO et l'IEC produisent à eux deux environ 85 % des standards internationaux (Büthe (T.), Mattli (W.), *The New Global Rulers: The Privatization of Regulation in the World Economy*, Princeton, Princeton University Press, 2013),

¹¹² Notamment leurs propriétés et caractéristiques.

Figure n° 2 : STRUCTURE DU ISO/IEC JTC 1/SC42

RÉFÉRENCE ↓	TITRE	TYPE
ISO/IEC JTC 1/SC 42/AG 2 ❶	Ingénierie des systèmes d'intelligence artificielle	Groupe de travail
ISO/IEC JTC 1/SC 42/AHG 1 ❷	Diffusion et sensibilisation	Groupe de travail
ISO/IEC JTC 1/SC 42/JWG 1 ❸	Groupe de travail mixte ISO/IEC JTC1/SC 42 - ISO/IEC JTC1/SC 40: Implications de l'IA en termes de gouvernance	Groupe de travail
ISO/IEC JTC 1/SC 42/WG 1 ❹	Normes fondamentales	Groupe de travail
ISO/IEC JTC 1/SC 42/WG 2 ❺	Données	Groupe de travail
ISO/IEC JTC 1/SC 42/WG 3 ❻	Fiabilité	Groupe de travail
ISO/IEC JTC 1/SC 42/WG 4 ❼	Cas d'utilisation et applications	Groupe de travail
ISO/IEC JTC 1/SC 42/WG 5 ❽	Caractéristiques systémiques et algorithmes pour la mise en œuvre de l'intelligence artificielle	Groupe de travail



Source : ISO

Il convient de préciser que les travaux du SC 42 sont encore en cours et que beaucoup reste à faire. Le traitement de certaines dimensions est ainsi bien plus avancé dans d'autres enceintes. Dans la recension qu'il fait des travaux du sous-comité, Thomas Zielke (qui représente l'Allemagne en son sein) explique ainsi que la compatibilité et l'interopérabilité ne sont pas encore véritablement au programme de travail de l'organisme et que les progrès dans ces domaines sont surtout liés aux initiatives émanant de la communauté de l'*open source*¹¹³.

¹¹³ Zielke (T.), « Is Artificial Intelligence Ready for Standardization? », in Yilmaz (M.), Niemann (J.), Clarke (P.), Messnarz (R.), eds., *Systems, Software and Services Process Improvement 27th European Conference, EuroSPI 2020, Düsseldorf, Germany, September 9–11, 2020, Proceedings*, Springer, 2020, pp. 259-274.

5.2. Traitement dans le cadre de l'*Institute of Electrical and Electronics Engineers (IEEE)*

Association professionnelle née en 1963, l'*Institute of Electrical and Electronics Engineers* est juridiquement une organisation à but non lucratif de droit américain. Structure internationale, elle compte plus de 400 000 membres de nationalités différentes (ingénieurs électriciens et électroniciens, informaticiens, professionnels du domaine des télécommunications...). Elle possède des branches dans plusieurs parties du monde¹¹⁴. L'organisation a pour but global de promouvoir la connaissance dans le domaine de l'ingénierie électrotechnique.

Outre ses fonctions d'organisation de conférences scientifiques et de publication de travaux scientifiques¹¹⁵, l'IEEE joue un rôle important dans l'établissement des normes internationales dans ces champs de compétences scientifiques et technologiques. Une branche dédiée de l'organisation, l'*IEEE Standards Association*, assure notamment les tâches de production de ses propres normes et de leur publication.

À la fin de l'année 2016, l'IEEE a publié un rapport intitulé *Ethically Aligned Design*¹¹⁶. Il s'agissait d'un document conséquent (plus de 130 pages) qui cherchait à traiter des questions éthiques liées au développement de l'Intelligence artificielle et des systèmes autonomes, et même à proposer des pistes pour le développement de cadres juridiques et de méthodologies pour concevoir des guides éthiques en matière de R&D.

Ce document – qui encourageait les parties prenantes à la conception et à l'utilisation des systèmes autonomes et dotés d'IA à prendre en compte prioritairement les facteurs liés à l'éthique dans leurs actions – était une version de travail. L'IEEE a ainsi proposé de le commenter (jusqu'à mars 2017), afin qu'une nouvelle version, éventuellement amendée puisse être écrite.

Cette étude était l'une des actions organisées dans le cadre de l'*IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems* (« Initiative mondiale IEEE »). Cette initiative, prise en charge par l'*IEEE Standards Association*, vise à s'assurer que les parties prenantes, en particulier les spécialistes techniques et technologiques, soient informées des questions éthiques liées à la conception et au déploiement des systèmes autonomes et intelligents et qu'ils soient formés à traiter prioritairement ces thématiques. L'objectif est de proposer des recommandations, provenant de pairs, pour que les produits, services et systèmes recourant à l'IA soient conformes à des normes éthiques définies en commun. Il s'agit donc d'une initiative intégrée visant à produire des principes, puis à s'assurer de leur transcription dans les pratiques.

¹¹⁴ Dans les 160 pays dans lesquels elle est représentée, chacune de ses branches prend la forme d'une société savante.

¹¹⁵ L'IEEE publie environ 170 journaux et magazines scientifiques et sponsorise en moyenne plus de 1 300 conférences annuellement dans le monde entier.

¹¹⁶ IEEE, *Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*, Version 1 – For Public Discussion, December 2016.

Figure n° 3 : PREMIÈRES SÉRIES DE STANDARDS DE L'IEEE



IEEE P7000™	Model Process for Addressing Ethical Concerns During System Design
IEEE P7001™	Transparency of Autonomous Systems
IEEE P7002™	Data Privacy Process
IEEE P7003™	Algorithmic Bias Considerations
IEEE P7004™	Standard on Child and Student Data Governance
IEEE P7005™	Standard on Employer Data Governance
IEEE P7006™	Standard on Personal Data AI Agent Working Group
IEEE P7007™	Ontological Standard for Ethically driven Robotics and Automation Systems
IEEE P7008™	Standard for Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems
IEEE P7009™	Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
IEEE P7010™	Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems

Source : IEEE

Cette initiative intègre également des actions de production de standards. Au sein de l'*IEEE Standards Association*, différents groupes de travail ont ainsi été créés. Plusieurs séries de standards (IEEE P7000™) sont en cours d'élaboration. Si certains d'entre eux traitent, de manière classique, de l'interopérabilité technologique, des fonctionnalités, de la sûreté et de la facilitation du commerce, d'autres abordent des questions spécifiques au croisement des considérations technologiques et éthiques.

L'effort de l'IEEE va même plus loin puisque l'organisation a également lancé l'*Ethics Certification Program for Autonomous and Intelligent Systems* (ECPAIS). Il s'agit de créer les spécifications pour que des dispositifs de certification et des processus de notation puissent être élaborés, en particulier dans les domaines de la transparence, de l'imputabilité et de la réduction des biais contenus dans les algorithmes appliqués aux systèmes autonomes et intelligents.

Insertion des acteurs américains dans les structures internationales

Les responsables politiques et les administrations des États-Unis ont investi la plupart des nombreuses initiatives internationales qui servent à créer des cadres globaux pour le développement de l'IA et pour le stimuler. Ils cherchent également à mettre en place des partenariats avec certains de leurs alliés et partenaires. Ces partenariats concernent souvent des actions communes scientifiques et technologiques.

Dans les discours des autorités et les commentaires proposés par les observateurs et les journalistes, cette dimension internationale de la politique américaine est souvent mise en rapport avec la montée en puissance de la Chine dans le domaine de l'IA¹¹⁷. Pékin a en effet affiché l'objectif d'être, à l'horizon 2030, la nation la plus puissante dans le domaine de l'Intelligence artificielle (R&D comme applications). Les efforts consentis pour réaliser cet objectif (en particulier les investissements publics) sont mis en avant dans de nombreux discours américains pour attester de l'existence d'une véritable compétition internationale et d'un risque au niveau stratégique.

La montée en puissance chinoise est bien évidemment décrite comme dangereuse du fait des applications militaires investiguées et du renforcement de la compétitivité économique du pays. Plus fondamentalement, les dirigeants américains dénoncent le fait que Pékin veut utiliser l'Intelligence artificielle pour asseoir son régime autoritaire, en particulier pour assurer un contrôle social et géographique des populations, et empêcher l'organisation et l'expression des mouvements de contestation (indépendantistes, démocratiques...). Les discours américains sont assez largement fondés sur l'idée d'un « choc des cultures », qui mettrait face à face un ensemble de pays – largement occidentaux – démocratiques et la Chine. Dans nombre de documents institutionnels et de positions officielles, il est donc affirmé que les États-Unis doivent mobiliser leurs alliés et partenaires pour parvenir à créer des cadres internationaux pour l'IA protecteurs des valeurs démocratiques.

¹¹⁷ Voir par exemple la tribune de Michael Kratsios, *Chief Technology Officer* de la Maison blanche, en mai 2020, dans le *Wall Street Journal* : « Artificial Intelligence Can Serve Democracy », May 27, 2020.

1. Positionnement dans le cadre des travaux de la Convention sur certaines armes conventionnelles

La Directive 3000.09 a assez largement constitué le socle du positionnement initial américain lors des débats du Groupe d'experts gouvernementaux (voir 2^{ème} Partie, § 2.1.3.). Ce document intègre notamment une définition des systèmes d'armes autonomes (« *a weapon that, once activated, can select and engage targets without further intervention by a human operator* »). Elle inclut également une distinction entre deux catégories dénommées systèmes d'armes « semi-autonomes » et « autonomes »¹¹⁸. L'utilité de cette séparation a été mise en cause par certains acteurs, car la sélection des cibles – définie peu précisément dans la Directive – est le seul marqueur pour les distinguer (et non la décision d'engagement)¹¹⁹. Certains observateurs considèrent ainsi que la référence américaine crée une « distinction sans différence »¹²⁰.

Les thématiques traitées dans la Directive correspondent toutefois globalement aux questions soulevées par les parties prenantes aux débats au sein de la Convention (notamment celles des « anti »-SALA). De même, les obligations générales que le document américain crée (voir 1^{ère} Partie, § 3.1.1.) répondent à certaines des demandes de ces mêmes acteurs. En dépit des problèmes définitionnels, la campagne « *Stop Killer Robots* » a ainsi initialement proposé un avis relativement favorable sur la Directive, considérant qu'il s'agissait de la première prise de position politique officielle sur les SALA de la part d'un État¹²¹. *Human Rights Watch*, qui coordonne le mouvement, a même qualifié cette politique de « *world's first moratorium on lethal fully autonomous weapons* », notamment parce qu'elle dispose d'une date d'expiration de dix ans et du fait des restrictions que la Directive applique à ce qu'elle définit comme des systèmes autonomes usant de la force létale.

En plus de la publication relativement précoce de la Directive, les États-Unis ont décidé de participer aux débats qui se sont progressivement organisés dans le cadre des réunions du groupe d'experts gouvernementaux. Lors de la rencontre inaugurale de ce dernier (novembre 2017), ils ont ainsi été parmi les premiers à produire des documents de travail. Ils ont proposé deux des 10 présentations écrites ayant servi à lancer les discussions¹²².

En fait, comme l'on fait remarquer certains observateurs, la durée d'exécution de 10 ans de la Directive 3000.09 est en réalité une pratique habituelle pour ce type de document¹²³. Le moratoire que saluait *Human Rights Watch* semble donc surtout être la conséquence d'un

¹¹⁸ Les systèmes semi-autonomes peuvent « employer de l'autonomie » pour un large portefeuille de « fonctions liées aux engagements » pour des « cibles individuelles ou des groupes de cibles spécifiques qui peuvent avoir été sélectionnés par des opérateurs humains ».

¹¹⁹ Gubrud (M.), « Semi-Autonomous and on Their Own: Killer Robots in Plato's Cave », *Bulletin of the Atomic Scientists*, 12 April 2015.

¹²⁰ Roff (H.), « Autonomous or 'Semi' Autonomous Weapons? A Distinction without Difference », *Huffington Post blog*, 16 January 2015.

¹²¹ Campaign to Stop Killer Robots, « New US Policy », 16 April 2013.

¹²² *Autonomy in Weapon Systems*, CCW/GGE.1/2017/WP.6 et *Characteristics of Lethal Autonomous Weapons Systems*, CCW/GGE.1/2017/WP.7.

¹²³ Bode (I.), Huelss (H.), « Autonomous weapons systems and changing norms in international relations », *Review of International Studies*, 44 (3), July 2018, pp. 393-413.

« réflexe » politico-administratif et non d'un acte volontaire réfléchi. Surtout, si le document interdit l'emploi des systèmes totalement autonomes, il pose également les contours des voies que pourraient employer les armées pour ne pas avoir à appliquer les restrictions que pourtant il établit. De même, le document précise que les systèmes d'armes semi-autonomes, définis assez largement, peuvent être employés pour appliquer des attaques létales comme non-létales.

Le point de vue de la campagne « *Stop Killer Robots* » vis-à-vis de l'approche des États-Unis a d'ailleurs fortement évolué. En plus des problèmes que pose la Directive 3000.09¹²⁴, principal document national de cadrage, le pays – comme d'autres (la Russie et Israël) – s'est officiellement positionné contre la création d'un nouveau traité d'*arms control*. Il s'est même également prononcé contre une déclaration politique commune contraignante. L'un des principaux arguments employés pour justifier cette position est qu'il est trop tôt pour construire un cadre trop rigide. Alors que l'on ne sait pas véritablement encore comment l'Intelligence artificielle va évoluer et sera utilisée par les armées dans le futur, il ne serait pas possible de produire un cadre de régulation véritablement adapté¹²⁵. Les États-Unis ont de plus fait valoir que 1/ le droit international humanitaire existant est, au stade actuel de développement de l'IA militaire et encore probablement pour plusieurs années, suffisant pour encadrer l'emploi des systèmes autonomes armés ; 2/ les procédures internes à chaque pays d'audit des systèmes d'armes sont essentielles ; 3/ les armements autonomes pourraient avoir des bénéfices humanitaires, notamment renforcer le respect des normes internationales¹²⁶.

Un autre point particulier distinguant le positionnement des États-Unis par rapport à ceux de nombreux autres acteurs des débats est qu'ils ne recourent pas au concept de « contrôle humain ». Ils considèrent en effet que cette notion obscurcit les réflexions plus qu'elle ne les éclaire. Ils utilisent ainsi les expressions « *human-machine interaction* » et « *appropriate levels of human judgement* », qui posent pourtant – comme nous l'avons déjà expliqué (1^{ère} Partie – § 3.1.1.) – des problèmes.

Les États-Unis considérant que les cadres juridiques internationaux existants étant suffisants, leur approche repose avant tout sur la conformité aux principes fondamentaux du droit international, qui doivent en permanence intégrer la spécificité des effets militaires¹²⁷. Cette approche ne tient pas véritablement compte du niveau de délégation de la prise de décision de l'humain vers la machine lors de la phase de sélection des cibles. Les États-Unis autorisent l'autonomie en matière de ciblage tant qu'il n'y a pas d'obligation légale obligeant à ce que le

¹²⁴ Notamment la distinction systèmes autonomes/systèmes semi-autonomes.

¹²⁵ Argument notamment présent dans United States, *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, working paper for the Group of Governmental Experts (2018-08-27–31), 2018.

¹²⁶ L'Intelligence artificielle pourrait notamment conférer aux systèmes de plus grandes performances que les êtres humains dans l'identification des civils et la production d'un jugement sur la façon d'utiliser la force de manière proportionnelle.

Pour un accès à la déclaration américaine sur ce point : <https://geneva.usmission.gov/2018/04/09/ccw-u-s-opening-statement-at-the-group-of-governmental-experts-meeting-on-lethal-autonomous-weapons-systems/>

¹²⁷ United States, *Human-Machine Interaction in the Development, Deployment and Use of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, working paper for the Group of Governmental Experts (2018-08-27–31), 2018.

Les États-Unis considèrent donc que l'autonomie militaire doit disposer d'un cadre juridique spécifique, différent de celui de l'autonomie pour les applications civiles, afin de ne pas empêcher les opérations des forces armées.

système d'arme lui-même soit programmé pour réaliser des évaluations en fonction du droit humanitaire¹²⁸.

2. Une recherche de partenaires privilégiés

Les documents institutionnels américains – notamment l'Initiative nationale et la Stratégie du DoD – appellent à une coopération internationale concernant la conception et l'utilisation de l'Intelligence artificielle, y compris dans le domaine de la défense¹²⁹. Cette demande, qui concerne les alliés et partenaires des États-Unis, a été reprise par nombre d'experts. Elle a reçu des échos chez certains d'entre eux. La Stratégie nationale en IA allemande, sortie en 2019, demande par exemple à ce que le pays travaille avec les nations *leaders* en Intelligence artificielle pour conduire des actions communes bilatérales ou multilatérales.

2.1. Relations bilatérales

2.1.1. « Relation spéciale » et IA

Lors du Sommet national d'août 2018, l'administration Trump avait indiqué que les priorités américaines étaient globales, ne concernant pas que le pays. À cette occasion, fut plus particulièrement mis en avant que des accords étaient en négociation, par l'*Office of Science and Technology Policy* de la Maison blanche, avec certains pays, dont le Royaume-Uni.

Ces discussions ont abouti à la signature d'un partenariat, en septembre 2020, entre les États-Unis, la Grande-Bretagne et l'Irlande du Nord. Ce document prend la forme d'une déclaration commune de coopération scientifique et technologique. Il ne précise cependant pas les formes que va emprunter cette coopération. Des discussions sont en cours à ce sujet, qui devraient continuer au moins jusqu'à la fin de la séquence électorale américaine. Les premiers éléments révélés indiquent toutefois que la collaboration entre chercheurs sera renforcée, de même que les échanges d'étudiants. Des partenariats public-privé bilatéraux pourraient également être pratiqués.

Les discussions actuelles concernent aussi la détermination des priorités en matière de R&D. La coopération doit cependant porter sur les actions de recherche pluridisciplinaire et sur les systèmes intelligents. Les commentaires de membres de l'Administration américaine ont notamment fait état de la volonté commune de travailler, de manière collaborative, sur les questions de l'explicabilité et de l'impartialité des systèmes d'IA¹³⁰.

¹²⁸ United States, *Implementing International Humanitarian Law in the Use of Autonomy in Weapon Systems*, working paper for the Group of Governmental Experts (2019-03-25–29), CCW/GGE.1/2019/WP.5, 2019.

¹²⁹ Dans l'un de ses premiers rapports, l'AI Commission résumait ce besoin de collaboration internationale : « *If the United States and its allies do not coordinate early and often on AI-enabled capabilities, the effectiveness of our military coalitions will suffer* » (National Security Commission on Artificial Intelligence, *Interim Report*, November 2019, p. 45).

¹³⁰ Gold (A.), « Exclusive: U.S. and U.K. announce AI partnership », Axios, September 25, 2020.

La coopération va être construite à partir du socle politique et juridique qui a été créé avec le *US-UK Science and Technology Agreement*. Signé en septembre 2017, celui-ci repose notamment sur un *Memorandum of Understanding* liant les organisations publiques civiles de recherche les plus puissantes des deux pays (*U.S. National Science Foundation* et *UK Research and Innovation*).

Alors que l'Union européenne avait publié en 2019 sa stratégie pour l'Intelligence artificielle, la signature de cet accord a été interprétée par certains observateurs comme une marque supplémentaire de la rupture entre l'UE et la Grande-Bretagne, et le choix fait par celle-ci d'affermir sa « relation spéciale » avec son allié américain.

Ce partenariat a été présenté par un certain nombre d'observateurs comme relevant de la volonté des États-Unis de mobiliser ses alliés et partenaires afin de contrer la montée en puissance chinoise dans le domaine de l'IA¹³¹. Dans sa présentation officielle du partenariat, Michael Kratsios, *U.S. Chief technology officer*, a ainsi rappelé qu'il devait servir à promouvoir l'innovation en IA pour le bien-être des citoyens et dans le respect des valeurs démocratiques partagées.

La recherche d'un partenariat consacré à l'IA entre les deux pays a également concerné les activités de défense. En mai 2018, des représentants du *Defense Innovation Board* américain ont ainsi rencontré leurs homologues britanniques, ainsi que les membres de l'*AI Lab*¹³². À cette occasion, les deux organismes ont déterminé les principaux domaines dans lesquels la coopération entre États-Unis et Grande-Bretagne allait se développer. À l'issue de la réunion, il a ainsi été précisé que les deux pays voulaient notamment échanger sur les thèmes de l'Intelligence artificielle, des véhicules autonomes, des capacités cyber avancées et des robots, aussi bien au niveau opérationnel que sur les thématiques strictement technologiques. D'autres visites ont par la suite été organisées¹³³.

Les échanges se sont accompagnés du lancement de projets communs. Le Dstl et l'*Engineering and Physical Sciences Research Council* (EPSRC), principale agence publique britannique de financement des travaux de recherche en ingénierie et en sciences physiques, soutiennent ainsi conjointement 4 programmes, en collaboration avec le DoD. Ces actions sont réalisées dans le cadre de l'initiative américaine MURI (*Multidisciplinary University Initiative*). Celle-ci a pour fonction de permettre l'étude, notamment grâce aux contributions d'équipes universitaires, de sujets prioritaires dont le traitement nécessite systématiquement les apports de plusieurs disciplines.

¹³¹ *Ibid.*

¹³² En mai 2018, le *Ministry of Defence* a annoncé la constitution, au sein du *Defence Science and Technology Laboratory* (Dstl), d'un *AI Lab*. Sa fonction est de coordonner les programmes de R&D de la Défense portant sur l'Intelligence artificielle, le *Machine learning* et les *data sciences*. Au moment de sa création, le Dstl gérait plus de 20 millions de Livres de budget attribués aux recherches en IA. Une grande partie de ces fonds a donc été transférée à l'*AI Lab*. Ces financements doivent augmenter à l'avenir.

Le *Lab* a pour fonction d'administrer les programmes et projets de recherche en IA conduits par le Dstl. Lors de l'annonce de la naissance de la structure, le Secrétaire à la Défense a précisé que, parmi les thématiques qu'elle allait plus particulièrement traiter, se trouvent les véhicules autonomes, la lutte contre les *fake news*, ainsi que le développement des cyber-protections avancées pour les réseaux informatiques. Le *Lab* doit également gérer la diffusion des savoir-faire en Intelligence artificielle au sein du ministère, notamment aider aux actions de sensibilisation et de formation. Son activité implique donc qu'il travaille étroitement avec les structures qui ont été mises en place au sein des armées et des services interarmées pour conduire les projets et programmes en IA.

¹³³ Plus tard dans l'année, le *Defence Innovation Board* britannique s'est ainsi rendu aux États-Unis.

Les initiatives communes avec les acteurs américains ne concernent pas que des programmes de recherche. En mars 2019, le Dstl et l'*U.S. Air Force Research Laboratory* (AFRL) ont organisé conjointement un hackathon concernant le *swarming* de drones. Il s'agissait pour les équipes engagées de développer des nouvelles méthodes pour utiliser des UAS pour assister les opérations de lutte contre les incendies¹³⁴.

2.1.2. Partenariat avec Singapour

En juin 2019, les États-Unis ont annoncé le début d'un nouveau partenariat en matière d'Intelligence artificielle avec Singapour¹³⁵. Cet accord porte sur le développement et les emplois des technologies d'IA dans le domaine de la sécurité nationale. Il repose notamment sur une coordination au travers d'échanges entre le JAIC américain et la *Defense Science and Technology Agency* singapourienne.

L'accord stipule que la coopération entre les deux pays doit essentiellement concerner les dispositifs apportant une aide pour l'assistance humanitaire régionale et les opérations de gestion des désastres (tremblements de terre). Le fait qu'il porte essentiellement sur des thématiques de sécurité nationale non strictement militaires peut s'expliquer par la volonté des dirigeants politiques d'éviter, notamment vis-à-vis de l'opinion publique, d'aborder des questions polémiques¹³⁶.

L'amplification des relations américaines, en particulier en matière de défense, avec Singapour est l'une des réponses apportées par Washington au comportement plus engagé de Pékin au niveau régional, ainsi qu'à la compétition qui s'affirme entre les deux pays¹³⁷.

Il est intéressant de noter que, pour la mise en place des partenariats avec Singapour et le Royaume-Uni, le choix a été fait de lancer certains travaux communs dans des domaines relativement consensuels : aide humanitaire, réponse aux catastrophes naturelles, lutte contre les incendies et cybersécurité. Une attitude équivalente avait été adoptée lorsque le JAIC a décidé, en application de la *DoD Artificial Intelligence Strategy*, d'initier ses premières *National Mission Initiatives*¹³⁸. Deux actions pilotes ont été lancées peu après l'adoption du document, qui ont concentré une grande partie de l'activité et des moyens du Centre¹³⁹. Il s'agissait de programmes devant permettre d'identifier des solutions existantes et de les adapter aux

¹³⁴ « Dstl Announce Swarming Drone Hackathon Challenge », MoD/Dstl press release, 13 February 2019.

¹³⁵ Parameswaran (P.), « What's in the New US-Singapore Artificial Intelligence Defense Partnership? » *The Diplomat*, July 1, 2019.

¹³⁶ Lin-Greenberg (E.), « Allies and Artificial Intelligence: Obstacles to Operations and Decision-Making », *Texas National Security Review*, 3 (2), Spring 2020, p. 72.

¹³⁷ Cortez (A. C.), Chase (M. S.), *Regional Responses to U.S.-China Competition in the Indo-Pacific*, Rand Corporation, 2020.

¹³⁸ Dans le cadre de la Stratégie du DoD, deux types de programmes ont été distingués : les *National Mission Initiatives* et les *Component Mission Initiatives*. Les premiers concernent des applications communes à toutes les armées. Il s'agit de projets portant sur des premiers cas d'usage, correspondant aux thématiques listées par la Stratégie. Ils sont directement administrés par le JAIC. Les seconds sont spécifiques à l'un des *Services* ou des *Combatant Commands*. Dans le cadre d'une *Component Mission Initiative*, le JAIC ne fait qu'assister, la gestion étant assurée par la composante intéressée. Ces deux types d'actions sont conçues comme complémentaires.

¹³⁹ Cronk (T. M.), « DOD Unveils Its Artificial Intelligence Strategy », U.S. Department of Defense, February 12, 2019.

besoins de la Défense (afin que des prototypes puissent être produits relativement rapidement – quelques mois) dans les domaines de la maintenance prévisionnelle et de l'assistance humanitaire et le secours aux sinistrés¹⁴⁰.

Ces thématiques suscitent en effet moins de polémiques et de réticences et permettent de montrer que l'IA utilisée pour la sécurité nationale – conçue assez largement – peut parfaitement participer au bien-être des populations. Elles sont également moins susceptibles de blocages du fait d'échanges de données sensibles et de différences culturelles (notamment morales) entre les partenaires. Il s'agit donc d'initier le partenariat, de rôder les mécanismes de coopération, avant d'intégrer des types d'applications plus complexes à aborder d'un point de vue éthique (mais qui sont prévus dans les accords).

2.2. Création de partenariats multilatéraux dédiés

Le DoD a lancé un partenariat sur les pratiques en matière d'Intelligence artificielle (« *AI Partnership for Defense* »). En septembre 2020, différents alliés et partenaires, européens, asiatiques et du Moyen-Orient, ont accepté d'y participer. Des délégations de 13 nations étaient ainsi présentes à la première réunion¹⁴¹. Organisée par le JAIC, elle s'est étalée sur deux journées et a notamment servi à partager les retours d'expériences nationaux et les meilleures pratiques. L'objectif principal était toutefois de faire réfléchir les représentations à ce que devrait être une « IA responsable »¹⁴².

Le partenariat, qui doit se poursuivre au travers d'autres réunions, est aussi prévu pour traiter de la question de l'interopérabilité, en particulier pour le partage des données, les ontologies et le développement des éléments permettant de rendre les systèmes d'IA sûrs.

Mark Beall, le *Chief of strategy* du JAIC, a ainsi indiqué que le partenariat a permis au DoD de présenter aux autres nations ces Principes éthiques en matière d'IA (voir 1^{ère} Partie, § 3.1.2.)¹⁴³. Il a cependant été précisé qu'il ne s'agissait pas de faire adhérer à ces principes les alliés et partenaires, mais de leur présenter la méthode de constitution employée afin de les aider à créer leurs propres références éthiques. Les techniques permettant de mettre en œuvre les Principes dans la pratique ont également été l'un des thèmes abordés au cours des deux jours.

Les organisateurs ont par ailleurs indiqué que, s'ils espéraient que d'autres pays se joignent au Partenariat à l'avenir, celui-ci a vocation à n'accueillir que des partenaires et alliés, c'est-à-

¹⁴⁰ En particulier l'exploitation en temps réel des données issues des capteurs pour la détection et le suivi de différents types de catastrophes (feux de forêt, tremblements de terre...). Il s'agissait notamment de capitaliser sur les premiers résultats du projet MAVEN en matière d'exploitation des données image et vidéo.

¹⁴¹ Australie, Canada, Danemark, Estonie, Finlande, France, Israël, Japon, Norvège, République de Corée, Suède, Royaume-Uni et États-Unis.

¹⁴² Tucker (P.), « France, Israel, S. Korea, Japan, Others Join Pentagon's AI Partnership », DefenseOne, September 19, 2020.

¹⁴³ *Ibid.*

dire des nations partageant des « intérêts mutuels » avec les États-Unis (des objectifs de politiques étrangères proches, ainsi qu'une adhésion aux conventions sur les droits de l'homme et le respect de l'état de droit)¹⁴⁴.

2.3. Un effort de mobilisation dans le cadre de l'OTAN

En janvier 2020, le G^{al} Shanahan, directeur du JAIC, a affirmé dans une interview téléphonique que les alliés devaient se joindre à l'effort américain de développement des capacités de l'IA militaire¹⁴⁵. Comme dans bien d'autres occasions, cette nécessité a été justifiée par le besoin d'acquérir une avance sur la Russie et la Chine. Les deux pays ont été décrits comme travaillant à la montée en puissance de l'Intelligence artificielle dans des voies contraires aux valeurs occidentales partagées.

Selon le G^{al} Shanahan, les autorités chinoises utiliseraient notamment les capacités étatiques de censure pour faciliter les ventes à l'étranger de systèmes d'armes intégrant de l'IA. Cette dissémination pourrait aboutir à doter des acteurs non-étatiques de capacités significatives. Il a également indiqué que les firmes de haute technologie chinoises, comme *Huawei*, étaient contraintes de coopérer avec les services de renseignement du régime. Il a complété ses propos en affirmant que la volonté de la Russie de ne pas respecter les normes éthiques internationales était manifeste et que le pays développe des systèmes pouvant avoir pour conséquence de déstabiliser la sécurité internationale (en particulier des outils pouvant amplifier les campagnes globales de désinformation).

Le plus haut responsable du JAIC a indiqué que, tandis que le Pentagone avait accéléré le rythme de développement de l'IA militaire, certains alliés progressaient lentement du fait d'une mobilisation sur les débats éthiques. Il a donc précisé que les actions communes pour l'adoption d'une IA responsable devraient assurer que les armées des membres de l'Alliance puissent suivre le rythme de la digitalisation et demeurer interopérables.

3. Des cadres multilatéraux investis avec prudence

Bien qu'elle ait été rédigée avant que les Principes de l'OCDE ne soient publiés (mai 2019), l'*American Artificial Intelligence Initiative* de février 2019 est globalement relativement bien « alignée » avec eux. Il ne semble pas exister de contradictions majeures entre les deux textes. Certains responsables américains l'ont d'ailleurs publiquement relevé¹⁴⁶.

Les États-Unis ont en effet assez largement participé aux travaux menés au sein de l'OCDE pendant toute l'année 2018, qui ont abouti à la rédaction des Principes. Par la suite, ils ont

¹⁴⁴ Barnett (J.), « DoD kicks off AI discussions with allies », *fedscoop*, September 16, 2020.

¹⁴⁵ Everstine (B. W.), « DOD AI Leader Wants Closer Collaboration With NATO », *Air Force Magazine*, January 15, 2020.

¹⁴⁶ Voir, par exemple, la présentation de l'*Initiative* proposée par Lynne Parker, alors *Chief Technology Officer* de la Maison blanche, en juin 2019 sur le site de l'OCDE (<https://oecd.ai/wonk/the-american-ai-initiative-the-u-s-strategy-for-leadership-in-artificial-intelligence> – consultée le 30 octobre 2020).

soutenu les efforts visant à déterminer les méthodes pour les mettre en œuvre. Ils contribuent notamment à l'*AI Policy Observatory* dont s'est doté l'Organisation (voir 2^{ème} Partie, § 2.3.).

En dépit de cette participation aux efforts de l'OCDE, la Maison blanche a appelé, au début de l'année 2020, les législateurs et le secteur commercial, ainsi que les alliés européens, à éviter une surréglementation de l'Intelligence artificielle¹⁴⁷.

Les États-Unis ont notamment eu une attitude fluctuante à l'égard du *Global Partnership on Artificial Intelligence*, lancé par la France et le Canada (voir 2^{ème} Partie, § 2.5.). Ils n'ont en effet pas fait partie, au cours des années 2018-2019 (période durant laquelle le projet se structurait), des pays qui avaient manifesté un intérêt. Lors du sommet du G7 de Biarritz, la représentation américaine a cherché à bloquer l'initiative. Début 2020, la Maison blanche avait ainsi décidé de ne pas y participer¹⁴⁸. Certains responsables politiques ont affirmé qu'elle pourrait aboutir à la mise en place d'un carcan trop restrictif et qu'elle venait dupliquer le mouvement lancé au sein de l'OCDE. Alors que certains acteurs soutenaient pourtant le projet, une partie des autorités politiques américaines, notamment au sein de l'exécutif¹⁴⁹, a fait savoir qu'elles considéraient qu'il apparaissait trop bureaucratique et surtout que les *inputs* produits pourraient entraver les activités des groupes américains de haute technologie dans le domaine de l'IA.

Finalement, les États-Unis ont décidé en juin 2020 de faire partie des « fondateurs » du *Global Partnership on Artificial Intelligence*. Il est probable que, devant le succès progressif du projet (notamment le nombre de pays qui s'y sont ralliés), les autorités américaines aient estimé qu'il valait mieux l'intégrer. Surtout, les États-Unis ont obtenu des créateurs originels, la France et le Canada, que le Partenariat soit pris en charge par l'OCDE et se rapproche donc de l'initiative lancée au sein de celle-ci (voir 2^{ème} Partie, § 2.3.). Ce rapprochement des deux projets accentue en effet la dynamique globale et permet de renforcer l'unité des pays démocratiques demandée par les États-Unis. La tribune rédigée par Michael Kratsios, en mai 2020, dans le *Wall Street Journal*, qui annonce le lancement officiel du GPIA (voir introduction à la présente partie), est d'ailleurs éclairante¹⁵⁰ : le *Chief Technology Officer* de la Maison blanche affirme que ce sont en effet les principes démocratiques qui unissent les membres fondateurs du Partenariat global. Dans le même temps, il précisait toutefois que la nouvelle coalition ne serait pas un organisme de créations de standards, ni un comité déterminant des politiques communes¹⁵¹.

L'insertion des États-Unis au sein des cadres internationaux traitant de l'IA répond donc à plusieurs logiques (présentées en 1^{ère} Partie, § 4.). Tout d'abord, il s'agit de fédérer un grand nombre de pays contre des opposants non-démocratiques. C'est surtout la Chine qui est – très clairement – identifiée à un risque. De manière assez transparente, les autorités américaines cherchent à mobiliser contre elle. Les États-Unis sont même capables de revenir sur leur positionnement initial s'ils constatent que celui-ci est susceptible de diviser plus que d'unir. Dans

¹⁴⁷ Sarwari (K.), « The G7 wants to regulate artificial intelligence. Should the US get on board? », TechXplore, January 14, 2020.

¹⁴⁸ *Ibid.*

¹⁴⁹ Notamment Lynne Parker, *Deputy chief technology officer* à la Maison blanche.

¹⁵⁰ Kratsios (M.), « Artificial Intelligence Can Serve Democracy », *Wall Street Journal*, May 27, 2020.

¹⁵¹ O'Brien (M.), « US joins G7 artificial intelligence group to counter China », DefenseNews, May 29, 2020.

le même temps, la constitution de ces cadres ne doit pas aboutir à ce que les acteurs américains se voient imposer des normes contraires à leurs intérêts, en particulier économiques. Les dispositifs trop contraignants sont ainsi rejetés.

Cette approche n'a cependant rien d'original. Dans une certaine mesure, elle a également été adoptée dans le domaine de la 5G. Dans ce domaine, les États-Unis produisent des discours alarmistes, indiquant que les matériels produits par *Huawei* ne sont pas sécurisés et créent des dépendances et des vulnérabilités pour ceux qui s'en dotent. Ils cherchent à constituer des coalitions pour que des solutions technologiques et industrielles soient développées et pour que les partenaires et alliés ne travaillent pas avec les entreprises chinoises. Derrière les questions d'autonomie stratégique et de non-respect des valeurs et droits « démocratiques », existe également une volonté de protéger les entreprises domestiques intervenant dans ce secteur.

4. Dominer l'édification des standards techniques internationaux

De nombreux acteurs considèrent que la production de standards – techniques et organisationnels – peut constituer une partie de la solution aux problèmes éthiques que posent l'IA et l'IA de défense. Or, de nombreux pays ont décidé de participer à la création des normes internationales en matière d'IA, voire de diriger les efforts dans ce domaine. Si certaines particularités nationales déterminent ces processus, trois éléments semblent cependant globalement les encadrer¹⁵² :

- ➔ Les acteurs clés au niveau mondial – notamment les États-Unis et la Chine – considèrent tous que des standards internationaux doivent désormais être fixés ;
- ➔ Les efforts de production des standards internationaux étant récents et encore largement inaboutis, les standards nationaux demeurent pertinents : la création de ces derniers se poursuit activement ;
- ➔ La structure du marché en matière d'Intelligence artificielle incite les pays à s'assurer que les standards internationaux soient fortement alignés sur leurs propres standards.

4.1. Efforts chinois en matière de standardisation

La Chine a affiché officiellement l'objectif de développer des normes favorables à ses champions nationaux. Dans leur étude du principal document officiel en la matière, *l'Artificial Intelligence Standardization White Paper* (voir ci-après), Sacks, Triolo et Ding ont isolé plusieurs facteurs expliquant les conditions de sa production¹⁵³. Tout d'abord, le gouvernement chinois cherche à renforcer la compétitivité internationale des industries nationales de l'IA. Pour ce

¹⁵² Cihon (P.), *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*, Technical Report, Future of Humanity Institute, University of Oxford, April 2019.

¹⁵³ Ding (J.), Triolo (P.), Sacks (S.), « Chinese Interests Take a Big Seat at the AI Governance Table », *New America*, June 20, 2018.

faire, il assiste ces compagnies dans le développement de leurs droits de propriété intellectuelle, devenus une composante majeure des systèmes technologiques globaux. L'accentuation de la présence chinoise dans les associations internationales productrices de standards relève de cette logique de protection des droits de propriété des entreprises nationales les plus importantes.

Ensuite, depuis les années 2000, la politique économique nationale est de développer les capacités d'innovation du pays. Les efforts de standardisation internes permettent notamment de renforcer l'écosystème de l'innovation domestique en assurant l'interopérabilité entre les acteurs nationaux de l'IA et la création de méthodes et d'outils d'évaluations de la qualité des productions¹⁵⁴.

Enfin, la volonté de construire des jeux de standards relève de logiques dépassant les seules préoccupations économiques. Les efforts chinois de standardisation auraient également pour fonction de protéger la vie privée des consommateurs et d'assurer un développement sûr et sécurisé de l'IA. Des parties entières du Livre blanc sont ainsi consacrées aux effets potentiels de l'Intelligence artificielle sur la vie privée. Ces effets sont décrits et analysés relativement précisément. Pour Sacks, Triolo et Ding, ces paragraphes – qui sont cohérents avec d'autres documents institutionnels – attestent bien d'une prise en compte de ces dimensions par les autorités. Ce traitement reposerait sur le constat que certains marchés domestiques – potentiellement très importants – pourraient être impactés si les utilisateurs n'ont pas confiance dans les solutions proposées par les groupes nationaux (notamment si ces utilisateurs refusent des atteintes à leur vie privée réalisées par les entreprises pour des raisons commerciales).

Pour organiser les efforts dans le domaine de la standardisation, les autorités ont, dans un premier temps, demandé une revue des pratiques mises en œuvre par les industries domestiques. Ce travail, exécuté dans la première moitié de l'année 2017, a permis de faire un état des lieux interne au pays. Surtout, des *best practices* ont été identifiées.

Cette revue a abouti à la publication d'un document dédié, intitulé *Standardization of AI Helps Industry Development*. Il est venu soutenir le *Three-Year Action Plan for Promoting the Development of New Génération of AI Industry*, plan triennal rendu public en décembre 2017, qui vise à favoriser l'adoption de l'IA dans les secteurs industriels de l'économie (en particulier les secteurs manufacturiers). Il doit servir à favoriser l'intégration de l'IA dans ces branches économiques, notamment en évitant que les concepteurs et les acteurs chargés de l'implémentation des dispositifs ne produisent des solutions qui ne soient pas interopérables.

En janvier 2018, une nouvelle étape a été franchie par la publication de l'*Artificial Intelligence Standardization White Paper*¹⁵⁵. Ce document de 98 pages a été rédigé par le *China Electronics Standardization Institute* (CESI), à la demande et sous le contrôle de la Commission nationale de gestion de la standardisation du ministère de l'Industrie et des Technologies de l'informa-

¹⁵⁴ Des protocoles de mise en forme des données standardisés permettent par exemple de lutter contre l'insularisation des jeux de données.

¹⁵⁵ Han (M.), « China Aims to Get the Jump on AI Standardization », Synced, January 25, 2018.

tion (MIIT). Une trentaine d'institutions de recherche, de formation supérieure et d'entreprises intervenant en IA ont participé à sa rédaction (dont les BATX, l'Université de Pékin et l'Université Tsinghua).

Le document appelle à établir rapidement des standards techniques pour favoriser le développement de l'IA. Il décrit et analyse les standards techniques existants (chinois et étrangers) et précise des cadres pour concevoir ceux qui manquent. Les cadres proposés sont généraux, puisqu'ils ont été conçus pour permettre la création de standards « systémiques » (et donc éviter la production de standards « dispersés »). Le document en présente cependant pour plus de 200 technologies du domaine de l'AI.

Au moment de la sortie du Livre blanc, les autorités ont également annoncé la constitution du *National Artificial Intelligence Standardization Group* et d'un *Expert Advisory Group*¹⁵⁶. Ces deux organismes opèrent notamment une supervision et une coordination des efforts chinois de normalisation au sein des structures internationales. Ils doivent également produire une information sur les travaux de normalisation dans les pays étrangers aux acteurs et autorités chinois. Enfin, ils ont été chargés d'analyser les stratégies étrangères appliquées dans les structures internationales.

Les autorités avaient par ailleurs décidé, en 2017, de la création de l'*Artificial Intelligence Industry Alliance*. Rassemblant plus de 200 membres, cette structure fédérative doit favoriser la diffusion des informations et la mise en œuvre des planifications gouvernementales. C'est la *China Academy of Information and Communications Technology*, autorité de régulation rattachée au MIIT, qui a été choisie pour l'animer. Parmi d'autres, les missions de l'Alliance sont de réaliser des évaluations par branche industrielle, mais également des technologies, produits, services, méthodologies et standards de test.

Les acteurs chinois se sont donc organisés pour pouvoir produire des standards domestiques. Le *New Generation Artificial Intelligence Development Plan*, qui fonde la politique nationale dans le domaine, prévoit en effet que des efforts conséquents soient consacrés à l'élaboration des standards dans les secteurs qui guident le développement de l'IA. Au niveau international, le CESI participe activement aux travaux du sous-comité dédié à l'IA créé au sein de l'ISO (ISO/IEC.JTC 1/SC 42 – voir 2^{ème} Partie, § 5.)¹⁵⁷. La réunion inaugurale de ce sous-comité (novembre 2017) a d'ailleurs eu lieu à Beijing. Il s'agit donc bien d'une politique à deux étages : s'assurer d'une certaine uniformisation domestique, tout en cherchant à influencer la production internationale des standards – action favorable à des alliances économiques et technologiques stratégiques et des transferts technologiques avec l'extérieur limités mais susceptibles d'améliorer la base technologique et industrielle chinoise¹⁵⁸.

¹⁵⁶ Ibidem.

¹⁵⁷ Luo (Y.), « China's Framework of AI Standards Moves Ahead », *The National Law Review*, July 16, 2018.

¹⁵⁸ Certains experts notaient ainsi au moment du lancement des premiers travaux que près de la moitié des standards technologiques industriels employés par les entreprises chinoises dans les secteurs *hi-tech* « intelligents » n'était corrélée à aucun standard international (Wübbecke (J.), Meissner (M.), Zenglein (M. J.), Ives (J.), Conrad (B.), *Made in China 2025: The Making of a High-Tech Superpower and Consequences for Industrial countries*, MERICS Papers on China, 2, December 2016).

4.2. Efforts des États-Unis

Depuis de nombreuses années, les États-Unis mettent l'accent sur le rôle des standards en IA. Les documents institutionnels rappellent l'importance de leur création pour favoriser la pénétration de l'avance technologique et industrielle américaine au sein des écosystèmes innovants dans le monde entier. Le *National Artificial Intelligence Research and Development Strategic Plan*, publié en octobre 2016, appelait ainsi au développement des standards et des bancs de test pour les technologies de l'IA, ainsi que de méthodes pour mesurer les progrès dans ces domaines. Il identifiait ainsi 10 domaines de standardisation¹⁵⁹.

Pendant longtemps, les autorités n'ont toutefois pas précisé comment ces efforts devaient se manifester au niveau international. Ce constat peut sans doute s'expliquer par le fait que, contrairement à la logique fortement centralisée de la Chine, les responsables américains considèrent traditionnellement que c'est très largement aux acteurs économiques d'être eux-mêmes *leaders* dans le développement des standards¹⁶⁰.

L'*Executive Order* de février 2019 (*Maintaining American Leadership in AI* – voir 1^{ère} Partie, § 2.3.) fait cependant de l'encadrement de la production des standards internationaux en matière d'Intelligence artificielle une priorité pour le pays. Il confiait au *National Institute for Standards and Technology* la mission de créer l'ébauche d'un plan permettant d'identifier les corpus de standards sur lesquels les autorités devaient travailler. En août 2019, le NIST – en réponse à la demande de la Présidence – a donc publié un plan fédéral¹⁶¹.

La prise en compte des implications éthiques et sociétales du développement de l'Intelligence artificielle a également été intégrée à la stratégie américaine en matière de standards. Afin d'assurer que les systèmes en IA puissent être fiables et dignes de confiance, 58 co-auteurs de 30 organisations différentes ont par exemple publié, en avril 2020, un rapport sur les techniques permettant d'améliorer la vérifiabilité du développement de l'IA.

La stratégie américaine en matière de standards semble présenter des difficultés. En dépit des efforts de coordination par le NIST, certains auteurs ont alerté les autorités sur le fait que l'approche étatique de la Chine dans ce domaine pourrait se révéler plus efficace que celle du pays, qui continue d'être relativement décentralisée. Elle pourrait notamment être contre-productive pour assurer une interopérabilité entre des systèmes domestiques d'innovation souvent en concurrence. Surtout, elle serait insuffisante pour gérer le « dilemme du géant aveugle »¹⁶². Le gouvernement serait en effet un acteur capable de prendre des décisions fondamentales pour l'avenir du développement d'une technologie, notamment au travers des

¹⁵⁹ *Software engineering, performance, metrics, safety, usability, interoperability, security, privacy, traceability, domain-specific standards.*

¹⁶⁰ L'*American National Standards Institute* (ANSI), institution privée à but non-lucratif, joue notamment un rôle majeur de représentation des intérêts industriels américains dans les activités de création de standards internationales non-encadrées par des traités.

¹⁶¹ US Department of Commerce, National Institute for Standards and Technology, *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools.*

¹⁶² David (P. A.), « Some New Standards for the Economics of Standardization in the Information Age, » in Dasgupta (P.), Stoneman (P.), eds., *Economic Policy and Technological Performance*, Cambridge, Cambridge University Press, 1987, pp. 206–39.

processus de standardisation, alors qu'il n'est pas, dans le système économique américain, le mieux informé pour réaliser les choix.

Au niveau de la Défense, comme la Stratégie pour l'IA le précise, le JAIC est le principal organisme de coordination en matière de standardisation. Il s'agit de produire des cadres qui s'appliqueront à l'ensemble des technologies de l'IA (*Machine learning*, ontologies, méthodes d'ingénierie des connaissances...). L'objectif est bien évidemment d'assurer une interopérabilité entre les différents systèmes qui vont être élaborés au sein du DoD, mais également leur intégration aux solutions de *cloud computing* et de *Big data*¹⁶³. Pour le Centre, les efforts en matière de standardisation doivent aussi permettre de rapprocher des processus de conception souvent disparates.

Le JAIC travaille ainsi actuellement avec d'autres composantes du DoD, mais également avec des agences gouvernementales civiles. Il est également en contact avec les représentants d'organisations internationales : OTAN, ISO, *International Committee for Information Technology Standards*, *Advanced Technology Academic Research Center*. Les échanges portent notamment sur :

- ➔ Le soutien et la conduite des actions de R&D en IA ;
- ➔ L'engagement dans le développement de standards pour l'IA ;
- ➔ La fourniture et le déploiement de produits et de services fondés sur des standards ;
- ➔ Le développement et la mise en œuvre de politiques de soutien (dont des politiques de régulation si nécessaire).

La production des standards est un processus de long terme, nécessitant des itérations. Le JAIC est censé assurer ce processus, par sa fonction de coordination. Il convient toutefois de préciser que certains acteurs institutionnels américains ont tenté de créer les conditions pour que des standards de fait s'imposent. Un très grand nombre de campagnes d'évaluation et de challenges sont actuellement organisées aux États-Unis, notamment par la DARPA et l'IARPA. Les promoteurs de ces exercices cherchent à en faire des événements incontournables, au niveau international, pour les équipes développant des solutions d'IA. De la sorte, ils imposent leurs méthodes de métrologie et de validation et mettent en place les conditions pour que leurs normes en matière de certification s'imposent. Surtout, en faisant de ces événements des rendez-vous scientifiques et technologiques majeurs, ils structurent les problématiques sur lesquelles les chercheurs et les industriels du monde entier travaillent – y compris dans le domaine de la standardisation.

¹⁶³ Joint Artificial Intelligence Center, « Leading in AI Standards », AI in Defense, May 2020.

Annexe 1

ACTIVITÉS DE LA NATO-STO LANCÉES DANS LE CADRE DE LA THÉMATIQUE « INTELLIGENCE ARTIFICIELLE ET BIG DATA POUR LA PRISE DE DÉCISION MILITAIRE »

ACTIVITY	ACTIVITY TYPE	TITLE
AVT-334	RTG	CDT on Augmented Reality (AR) to Enhance Situational Awareness for Armored Fighting Vehicle Crew
AVT-353	RWS	Artificial Intelligence in Cockpits for UAVs
AVT-356	RSY	Physics of Failure for Military Platform Critical Subsystems
AVT-SP-007	SP	Novel Active Fire Modelling and Prediction methods using Manned and Unmanned Aircraft Vehicles
HFM-276	RTG	Human Factors and ISR Concept Development and Evaluation
HFM-281	RTG	Personalized Medicine in Mental Health and Performance
HFM-293	RTG	Digital And Social Media Assessment For Effective on Communication And Cyber Diplomacy
HFM-294	RTG	Big Data In The Military: Integrating Genomics into the Pipeline of Standard-care Testing & Treatment
HFM-305	RTG	Synthetic Biology in Defence: Opportunities and Threats
HFM-311	RTG	Cognitive Neuro-enhancement: Techniques and Technology
HFM-316	RTG	Expert panel for state of the art cardiovascular risk assessment in aircrew and other high-risk occupations.
HFM-317	LTSS	Solutions for Combat Casualty Care
HFM-322	RWS	Meaningful Human Control of AI-based Systems: Key Characteristics, Influencing Factors and Design Considerations
HFM-324	RSY	RSY; Solutions for Combat Casualty Care
HFM-330	RTG	Human Systems Integration for Meaningful Human Control over AI-based systems
IST-141	RTG	Exploratory Visual Analytics
IST-146	RTG	Electromagnetic Environment Situational Awareness for NATO
IST-157	RTG	Human in the Loop Considerations for Artificial Intelligence
IST-159	RTG	Cyber Intelligence and Social Media
IST-163	RTG	Deep Machine Learning for Cyber Defense

Observatoire de la politique de défense des États-Unis

Rapport 8 – Positionnement des acteurs américains sur les questions éthiques liées à l'intelligence artificielle de défense

IST-165	RTG	High-level Fusion Of Hard And Soft Information For Intelligence
IST-169	RTG	Robustness and Accountability in Machine Learning Systems
IST-171	RTG	FMN Cloud-based Coalition Security Architecture
IST-172	RTG	Airborne Beyond Line of Sight Communication Network
IST-177	RTG	Social Media Exploitation for Operations in the Information Environment
IST-178	RWS	Big Data Challenges: Situation Awareness and Decision Support
IST-183	RWS	Deep Machine Learning For Cyber Defense
IST-HFM-182	RTG	Synthetic Legal Adviser - AI-based Decision Making in HyperWar
MSG-155	RTG	Data Farming Services (DFS) for Analysis and Simulation-Based Decision Support
SAS-139	RTG	NATO Analytical War Gaming - Innovative Approaches for Data Capture, Analysis and Exploitation
SAS-143	RTG	Agile, Multi-Domain C2 of Socio-Technical Organizations in Complex Endeavors
SAS-150	ST	Advanced Analytics and Artificial Intelligence for Defence Enterprise Resource Planning
SAS-157	RTG	Automation in the Intelligence Cycle
SAS-158	RTG	Employing AI to Federate Sensors in Joint Settings
SAS-159	RTG	How could Technology Development Transform the Future Operational Environment
SAS-160	RTG	Ethical, legal and moral (ELM) impacts of novel technologies on NATO's operational advantage – the "ELM Tree"
SCI-331	ST	Fostering and Managing the STO Autonomy Portfolio
SCI-334	RTG	Evaluation of Swarm System for Military Applications
SCI-339	ST	Allied Future Surveillance & Control (AFSC) High Level Technical Concepts (HLTCs) Technology Readiness Level (TRL) Assessments
SET-245	RTG	Radar Based Non-Cooperative Target Recognition (NCTR) in the Low Airspace and Complex Surface Environments
SET-250	RTG	Multi-Dimensional Radar Imaging
SET-256	RTG	Interoperability & Networking of Disparate Sensors and Platforms for ISR Applications
SET-263	RTG	Swarms Systems for Intelligence Surveillance & Reconnaissance
SET-272	RTG	Automated Scene Understanding for Battlefield Awareness
SET-273	RSM	Multidimensional Radar Imaging and ATR
SET-277	RWS	RWS on Phenomenology and Exploitation of Hyperspectral Sensing within NATO
SET-278	RTG	Machine Learning for Wide Area Surveillance
SET-279	RTG	Space-based SAR and Big Data Technologies to support NATO Operations
SET-283	RTG	Advanced Machine Learning ATR using SAR/ISAR data
SET-286	RTG	Acoustic and Seismic Sensing of Threats in Urban Environments
SET-290	RLS	AI for Military ISR Decision Makers

Source : NATO Science and Technology Organization, *Collaborative Program of Work and budget for year 2020*.

Annexe 2

ACTIVITÉS DE LA NATO-STO LANCÉES DANS LE CADRE DE LA THÉMATIQUE « AUTONOMIE »

ACTIVITY	ACTIVITY TYPE	TITLE
AVT-310	RTG	Hybrid/Electric Aircraft Design and Standards , Research and Technology (HEADSTART)
AVT-336	RSM	Enabling Platform Technologies for Resilient Small Satellite Constellations for NATO Missions
AVT-341	RTG	Mobility Assessment Methods and Tools for Autonomous Military Ground Systems
AVT-353	RWS	Artificial Intelligence in Cockpits for UAVs
HFM-311	RTG	Cognitive Neuro-enhancement: Techniques and Technology
HFM-322	RWS	Meaningful Human Control of AI-based Systems: Key Characteristics, Influencing Factors and Design Considerations
HFM-330	RTG	Human Systems Integration for Meaningful Human Control over AI-based systems
IST-164	RTG	Securing Unmanned and Autonomous Vehicles for Mission Assurance
IST-174	RTG	Secure Underwater Communications for Heterogeneous Network-enabled Operations
IST-179	RTG	Interoperability for Semi-Autonomous Unmanned Ground Vehicles
IST-183	RWS	Deep Machine Learning For Cyber Defense
IST-186	RSM	Blockchain Technology for Coalition Operations
IST-HFM-182	RTG	Synthetic Legal Adviser - AI-based Decision Making in HyperWar
MSG-154	RTG	Low Slow Small Threats Modelling and Simulation
SAS-139	RTG	NATO Analytical War Gaming - Innovative Approaches for Data Capture, Analysis and Exploitation
SAS-143	RTG	Agile, Multi-Domain C2 of Socio-Technical Organizations in Complex Endeavors
SAS-152	RTG	Conceptual framework for Comprehensive National Defence System
SAS-157	RTG	Automation in the Intelligence Cycle
SAS-159	RTG	How could Technology Development Transform the Future Operational Environment
SAS-160	RTG	Ethical, legal and moral (ELM) impacts of novel technologies on NATO's operational advantage – the "ELM Tree"
SCI-301	RTG	Defeat of Low Slow and Small (LSS) Air Threats
SCI-321	RTG	UAV Applications for Military Search
SCI-324	RWS	Realization and Evaluation of Robotic Multispectral Decoys for Land Equipment
SCI-328	RSY	Flight Testing of Unmanned Aerial Systems (UAS)
SCI-330	RLS	Multinational Capability Development Campaign (MCDC) Analysis on Autonomous Unmanned Systems (UAxS)
SCI-331	ST	Fostering and Managing the STO Autonomy Portfolio
SCI-334	RTG	Evaluation of Swarm System for Military Applications
SCI-335	RSM	Autonomy from a System Perspective – Version 2.0
SCI-338	AG	Flight Testing of Unmanned Aerial Vehicles
SET-263	RTG	Swarms Systems for Intelligence Surveillance & Reconnaissance
SET-272	RTG	Automated Scene Understanding for Battlefield Awareness
SET-290	RLS	AI for Military ISR Decision Makers

Source : NATO Science and Technology Organization, *Collaborative Program of Work and budget for year 2020*.