ÉTUDES
DE L'IFRI

CENTRE GÉOPOLITIQUE DES TECHNOLOGIES

FÉVRIER
2025

Promesses artificielles ou régulation réelle ?

Inventer la gouvernance mondiale de l'IA

Laure de ROUCY-ROCHEGONDE

L'Ifri est, en France, le principal centre indépendant de recherche, d'information et de débat sur les grandes questions internationales. Créé en 1979 par Thierry de Montbrial, l'Ifri est une fondation reconnue d'utilité publique par décret du 16 novembre 2022. Elle n'est soumise à aucune tutelle administrative, définit librement ses activités et publie régulièrement ses travaux.

L'Ifri associe, au travers de ses études et de ses débats, dans une démarche interdisciplinaire, décideurs politiques et experts à l'échelle internationale.

ISBN : 979-10-373-0982-2

Couverture : Montage réalisé sur Canva par Emma Badaoui © Ifri, 2025

Comment citer cette publication :
Laure de Roucy-Rochegonde, « Promesses artificielles ou régulation réelle ? Inventer la gouvernance mondiale de l'IA », *Études de l'Ifri*, Ifri, Février 2025.

Ifri

27 rue de la Procession 75740 Paris Cedex 15 – FRANCE

Tél. : +33 (0)1 40 61 60 00 – Fax : +33 (0)1 40 61 60 60

E-mail : accueil@ifri.org

Site internet : Ifri.org

Autrice

Laure de Roucy-Rochegonde est directrice du Centre géopolitique des technologies de l'Ifri depuis février 2024. Elle était précédemment chercheuse au Centre des études de sécurité depuis 2019, où elle a travaillé sur les applications militaires de l'Intelligence artificielle, la conflictualité normative et la maîtrise des armements. Docteure en science politique, elle est également chercheuse associée au Centre de recherches internationales (CERI, Sciences Po/CNRS). En octobre 2024 est paru son premier ouvrage, intitulé *La Guerre à l'ère de l'Intelligence artificielle : quand les machines prennent les armes* (PUF).

Parallèlement à ses activités de recherche, elle enseigne l'éthique de la guerre et la maîtrise des armements à Sciences Po Paris et à l'université Paris 2 Panthéon-Assas. Elle est par ailleurs titulaire d'un master de politiques publiques et d'un *bachelor* de Sciences Po Paris, au cours duquel elle a passé une année au département de *War Studies* du King's College à Londres.

Résumé

Fruit des ambitions politiques et économiques d'une pluralité d'acteurs aux intérêts souvent divergents, l'encadrement international de l'Intelligence artificielle (IA) reflète avec acuité les tensions géopolitiques contemporaines.

Les risques inhérents au développement et à l'adoption massive de l'IA, technologie clé et vecteur de transformations profondes au sein des sociétés, pour la santé, l'éducation, l'emploi ou l'environnement, soulignent l'urgence d'harmoniser les efforts de gouvernance à l'échelle internationale.

La gouvernance mondiale de l'IA repose sur la capacité des acteurs étatiques et non étatiques à établir des normes communes sur les risques technologiques, les limites à établir, ainsi que les principes à garantir. Ces efforts visent à promouvoir un développement sécurisé de l'IA, universel, adapté aux diversités culturelles, exempt de biais, et respectueux des valeurs démocratiques ainsi que des droits et libertés fondamentaux.

Cependant, des défis politiques, économiques et juridiques résiduels exacerbés par les limites des cadres réglementaires existants – face à une balkanisation croissante des approches de gouvernance et à la fragmentation de la communauté internationale – complexifient considérablement la mise en œuvre d'une telle initiative.

Par ailleurs, compte tenu de la nature intrinsèquement évolutive de l'IA, il est indispensable que soit élaboré un cadre de gouvernance adaptable et flexible, « future-proof », à même d'anticiper les avancées techniques et de s'y ajuster.

La tenue du Sommet pour l'action sur l'Intelligence artificielle à Paris, début février, est une opportunité sans précédent de s'accorder sur une vision partagée de la gouvernance de l'IA, durable et inclusive. Pour les décideurs, c'est l'opportunité de mieux saisir l'évolution des pratiques, des insuffisances réglementaires, des intérêts qui influencent les accords en construction, et des compromis nécessaires pour encadrer l'IA à l'échelle mondiale dans les années à venir.

Executive summary

Arising from the political and economic ambitions of a plurality of players with often divergent interests, the international framework of artificial intelligence (AI) is an acute reflection of contemporary geopolitical tensions.

The risks inherent to the development and mass adoption of AI, a key technology and vector of profound transformations within societies for health, education, employment or the environment, underline the pressing need to harmonize governance efforts at the international level.

Global governance of AI relies on the ability of state and non-state players to set common standards on technological risks, the boundaries to be drawn, and the principles to be safeguarded. These endeavors aim to promote the safe development of AI that is universal, adapted to cultural diversities, free from bias, and respectful of democratic values and fundamental rights and freedoms.

However, residual political, economic and legal challenges exacerbated by the limits of existing regulatory frameworks - in the face of increasing balkanization of governance approaches and fragmentation of the international community - considerably complicate the implementation of such an initiative.

Given the intrinsically evolving nature of AI, it is vital to build an adaptable and flexible « future-proof » governance framework capable of anticipating and adjusting to technical advances.

The Summit for Action on Artificial Intelligence to be held in Paris in February is an unprecedented timely occasion to agree on a shared vision of AI governance that is sustainable and inclusive. For decision-makers, it's an opportunity to better grasp the evolution of practices, regulatory shortcomings, the interests influencing the agreements under construction, and the compromises needed to frame AI on a global scale in the years to come.

Sommaire

Introduction

Les 10 et 11 février 2025 se tiendra à Paris l'Artificial Intelligence Action Summit, qui réunira des chefs d'État et de gouvernement, des représentants d'organisations internationales, des dirigeants d'entreprises, des acteurs du monde universitaire, des organisations non gouvernementales (ONG), des artistes et des membres de la société civile venus de plus de 100 pays. Ce sommet, co-présidé par l'Inde, s'inscrit dans la continuité de l'Artificial Intelligence Safety Summit, organisé à Bletchley Park (Royaume-Uni) en novembre 2023 sous la houlette du gouvernement britannique – et qui a donné lieu à deux autres éditions, à Séoul en mai 2024 et à San Francisco en novembre 2024. L'infléchissement de l'approche par la France, qui décide de dédier son événement mondial à l'action plutôt qu'à la sécurité en matière d'Intelligence artificielle (IA), est emblématique de la tension entre risques et opportunités qui rend si difficile la régulation internationale du développement et des usages de cette technologie.

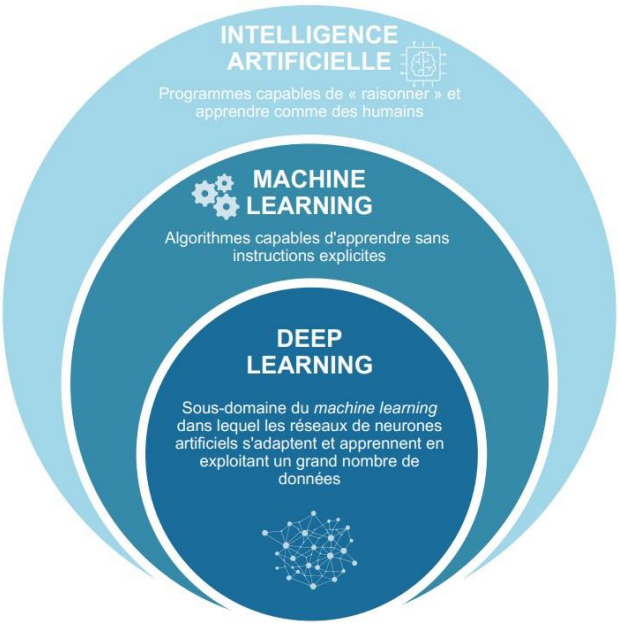
L'expression « Intelligence artificielle » a été inventée en 1956 par le logicien John McCarthy pour qualifier les techniques permettant de mieux appréhender l'intelligence humaine et de l'imiter grâce à des programmes informatiques[1]. Ainsi que l'indique le groupe d'experts de haut niveau de l'Union européenne (UE) sur l'IA constitué par la Commission européenne en juin 2018, il s'agit donc de « systèmes qui affichent un comportement intelligent en analysant leur environnement et en prenant des mesures – avec un certain degré d'autonomie – pour atteindre des objectifs spécifiques ».

Sont souvent distinguées schématiquement deux grandes familles de techniques d'IA :

- la modélisation des connaissances et du raisonnement (IA symbolique), qui consiste en une formalisation de connaissances sur lesquelles s'appuient des algorithmes pour produire des résultats ;
- les méthodes fondées sur les données, en particulier l'apprentissage (IA connexionniste), que l'on retrouve également dans les modèles d'IA générative[2].

1. J.-C. Noël, « Intelligence artificielle : vers une nouvelle révolution militaire ? », *Focus stratégique*, n° 84, Ifri, octobre 2018.
2. Cette dernière est le fruit des progrès en apprentissage profond (*deep learning*) et en traitement automatique du langage (*natural langage processing*), qui permettent à des systèmes informatiques de « saisir » le langage humain et d'accomplir des tâches complexes, notamment la génération automatisée de contenus textuels ou visuels à partir d'instructions (*prompts*).

Schéma des différentes techniques d'intelligence artificielle



Les techniques d'IA irriguent un nombre croissant de domaines de la vie quotidienne, et plus généralement de l'économie mondiale. En témoigne le succès commercial de ChatGPT, qui atteignait déjà les 100 millions d'utilisateurs quelques semaines après son lancement, à la fin de l'année 2022. Deux ans plus tard, l'agent conversationnel d'OpenAI dépasse désormais les 300 millions d'utilisateurs actifs hebdomadaires[3].

Depuis 2023, l'IA fait l'objet d'une couverture médiatique considérable. Si les espoirs qui y sont associés n'en finissent plus d'être mis en avant – amélioration des soins de santé, transports plus sûrs et plus écologiques, gains de productivité, énergie moins onéreuse et plus durable – les risques qui les accompagnent sont eux aussi régulièrement énoncés. Ainsi, en mai 2023, le « parrain de l'IA » Geoffrey Hinton – qui a reçu l'année dernière le prix Nobel de physique pour ses « découvertes et inventions fondamentales permettant l'apprentissage automatique au moyen de réseaux neuronaux artificiels » – démissionnait avec fracas de Google, alertant sur les dangers de son champ de recherche. Dans un entretien accordé au *New York Times*, il prévenait que « les futures versions de cette technologie pourraient être un risque pour l'humanité ». Il s'y inquiétait de la sophistication grandissante des techniques d'IA et du risque qu'elles ne tombent entre de mauvaises mains. « Une part de moi-même regrette l'œuvre de ma vie[4] », confiait-il aussi.

3. E. Roth, « ChatGPT Now Has over 300 Million Weekly Users », *The Verge*, 4 décembre 2024.
4. C. Metz, « The Godfather of AI Quits Google and Warns of Danger Ahead », *The New York Times*, 4 mai 2023.

L'IA est au cœur d'une véritable course à la puissance, dominée sans surprise par la rivalité sino-américaine mais où d'autres États, à l'image de la France, du Royaume-Uni, de l'Allemagne, de l'Inde, du Canada, de la Corée du Sud ou encore des pays du Golfe, tentent aussi de tirer leur épingle du jeu[5]. Dans cet environnement marqué par une compétition intense sur les plans politique, géopolitique et économique, les perspectives et les discours divergent quant aux priorités en matière de gouvernance de l'IA[6]. Tandis que certains mettent l'accent sur la nécessité de réguler, d'autres insistent sur le besoin de stimuler l'innovation et présentent les tentatives d'encadrement comme un obstacle aux découvertes. Derrière ces différences d'approche se dessinent des priorités nationales profondément divergentes, liées à une variété d'intérêts et de rapports aux normes.

Parce que les technologies d'IA s'inscrivent dans des chaînes de valeur transnationales et qu'elles tendent à se démocratiser et à se diffuser de manière exponentielle, leur réglementation ne peut être pensée qu'à l'échelle internationale. Pourtant, aux pierres d'achoppement classiques du multilatéralisme s'ajoutent les spécificités techniques de l'IA, qui rendent son encadrement d'autant plus délicat. Comment, alors, concevoir une gouvernance internationale de l'IA ?

La gouvernance globale, qui renvoie à « l'effort collectif d'États souverains, d'organisations internationales et d'autres acteurs non étatiques pour relever des défis communs et saisir des opportunités qui transcendent les frontières nationales[7] » est un débat classique en relations internationales. Au cœur de cette controverse se pose la question de la mise en œuvre d'une régulation universelle pour des enjeux de sécurité globaux.

Depuis 2019 – avec une accélération à partir de 2023 –, les initiatives pour encadrer l'IA se sont multipliées, portées par un large éventail d'acteurs, allant des gouvernements aux organisations internationales et régionales, en passant par des coalitions d'entreprises et des associations de la société civile. Fin octobre 2023, le G7 a ainsi adopté un code de conduite non contraignant destiné aux développeurs d'IA[8]. Quelques jours plus tard, en novembre, le Royaume-Uni organisait le Sommet de Bletchley Park ; tandis qu'après avoir promulgué un décret visant à promouvoir une IA « sûre, sécurisée et digne de confiance », le président américain Joe Biden rencontrait son homologue chinois Xi Jinping afin d'instaurer un dialogue bilatéral sur les usages militaires de l'IA. Début 2024, les législateurs européens trouvaient quant à eux un accord politique sur l'*AI Act*, un texte précurseur entré en vigueur en août de la même année, visant à encadrer les

5. B. Pajot, « L'Intelligence artificielle ou la course à la puissance », *Politique étrangère*, vol. 89, n° 3, Ifri, septembre 2024.
6. J. B. Bullock *et al.* (dir.), *The Oxford Handbook of AI Governance*, Oxford, Oxford University Press, 2022.
7. S. Patrick, « The Unruled World: The Case for Good Enough Global Governance », *Foreign Affairs*, vol. 93, n° 1, hiver 2014.
8. R. Balenieri, « IA : les pays du G7 adoptent un code de bonne conduite », *Les Échos*, 30 octobre 2023.

risques de cette technologie tout en établissant une référence mondiale en matière de régulation.

Ce paysage fragmenté risque toutefois d'engendrer des cadres de gouvernance disparates, des dialogues parallèles non coordonnés et des préférences collectives divergentes, qui pourraient compromettre l'innovation et entraver le développement de l'IA pour le bien commun. L'objectif de cette étude est alors de comprendre les blocages de la gouvernance de l'IA, afin de les surmonter. Dans un premier temps sera démontrée la nécessité d'une approche mondiale de l'encadrement de l'IA. Puis sera mise en lumière la « balkanisation » dont pâtit actuellement sa gouvernance. Enfin seront proposées des pistes pour une meilleure régulation internationale de l'IA.

Une préoccupation de gouvernance mondiale

Les avancées récentes des *Larges Langage Models* (LLM[9]) et des agents conversationnels tels que ChatGPT (OpenAI), Gemini (Google) ou Ernie Bot (Baidu) ont démocratisé le recours à l'IA générative. En se saisissant de ces outils, le grand public s'est familiarisé avec les multiples capacités de l'IA, suscitant dans le même mouvement un enthousiasme marqué et une montée des préoccupations.

À partir de 2023, chercheurs et responsables politiques ont accentué leurs mises en garde sur les dangers de cette technologie, notamment au regard des risques de suppressions d'emplois, des menaces pour la démocratie, des atteintes aux libertés civiles et à la vie privée, et des périls pour la propriété intellectuelle et le droit d'auteur[10]. L'impératif et l'urgence d'établir une réglementation afin que l'IA soit conçue et utilisée dans le respect du droit sont alors devenus des thèmes récurrents dans le débat public. Au besoin de préserver la sécurité nationale et les droits humains s'ajoute toutefois la nécessité de maintenir une compétitivité économique. Or, nombreux sont les acteurs qui pourfendent l'idée de se « lier les mains » en régulant une technologie que d'autres puissances pourraient utiliser sans limite.

L'IA, en raison de ses immenses répercussions économiques, politiques et sociales, soulève donc de nombreux défis en matière de gouvernance. Si ces enjeux étaient initialement pris en charge par les autorités nationales, les initiatives portées par des organisations internationales ont connu un essor considérable ces dernières années[11]. Cette première partie s'attache donc à démontrer pourquoi il est nécessaire d'encadrer le développement et le déploiement de l'IA, et pourquoi cette démarche n'a de sens qu'à l'échelle globale.

9. Les LLM désignent des modèles de fondation entraînés à l'aide d'immenses quantités de données pour comprendre et générer des textes en langage naturel, ainsi que d'autres types de contenu, afin d'accomplir un large éventail de tâches.
10. M. Schaake, « The Premature Quest for International AI Cooperation », *Foreign Affairs*, 21 décembre 2023.
11. *Ibid*.

Endiguer les risques de l'IA

Pourquoi l'IA est-elle devenue une priorité pour les régulateurs du monde entier ? La réponse à cette question se trouve probablement dans la mise en lumière des risques associés à cette technologie.

Des déclarations de Vladimir Poutine annonçant dès 2017 que « celui qui deviendra leader dans l'IA sera le maître du monde » à celles d'Elon Musk ajoutant que « l'IA est bien plus dangereuse que l'arme nucléaire » et « causera probablement une troisième guerre mondiale », les dernières années, en effet, ont vu les alertes sur les menaces sous-jacentes de l'IA se multiplier. Comme le souligne toutefois le philosophe et historien Émile Torres, ces prédictions apocalyptiques désormais convenues servent de leurre et détournent l'attention de problèmes pourtant bien réels :

> « Parler de l'extinction de l'humanité, d'un véritable événement apocalyptique, est tellement plus captivant que de parler des travailleuses et des travailleurs kényans payés 1,32 $ de l'heure pour modérer des contenus utilisés par l'IA, ou le travail d'artistes exploités pour alimenter ces systèmes.[12] »

Les risques de l'IA font en effet l'objet d'une compétition discursive[13]. Les géants du numérique exhortent à se pencher sur des menaces à long terme parfois farfelues (allant jusqu'à la peur de l'extinction de l'espèce humaine[14]) pour éclipser les dangers plus immédiats et tangibles (en ce qui concerne la propriété intellectuelle ou la fiscalité par exemple). Il n'en demeure pas moins que les revendications d'encadrement résultent très directement de la prise en compte de ces aléas. Quels sont, alors, les principaux risques associés à l'IA qui nécessiteraient une régulation ?

Pour le « bien commun »

Le développement fulgurant de l'IA soulève d'abord de multiples préoccupations sur la préservation du « bien commun », en lien avec des enjeux éthiques, sociaux, économiques et environnementaux[15].

Premièrement, comme le souligne la rapporteuse spéciale de l'Organisation des Nations unies (ONU) sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et d'intolérance,

12. Les travaux d'Émile Torres alertent sur la manière dont les concepts de « sûreté de l'IA » et de « recherche du bien de l'humanité » permettent en réalité à des courants de pensée d'influer sur les priorités de développement en IA et de se soustraire aux obligations de rendre des comptes au plus grand nombre. Lire É. Torres, *Human Extinction: A History of the Science and Ethics of Annihilation*, Londres, Routledge, 2023.
13. B. Pajot, « Les risques de l'IA : enjeux discursifs d'une technologie stratégique », *Études de l'Ifri*, Ifri, juin 2024.
14. K. Roose, « A.I. Poses "Risk of Extinction", Industry Leaders Warn », *The New York Times*, 30 mai 2023.
15. M. Coeckelbergh, « Artificial Intelligence, the Common Good, and the Democratic Deficit in AI Governance », *AI and Ethics*, 2024.

Ashwini K.P., « les récentes avancées dans le domaine de l'IA générative et l'essor des applications de l'IA continuent de soulever d'importantes questions relatives aux droits humains, notamment en ce qui concerne la discrimination raciale[16] ». Les systèmes d'IA tendent en effet à reproduire et à amplifier les biais présents dans les corpus d'entraînement, parce que les jeux de données utilisés pour entraîner les algorithmes sont généralement incomplets, et que certains groupes de personnes y sont sous-représentés[17]. Or, la surreprésentation ou la sous-représentation de groupes particuliers dans les jeux de données d'apprentissage, notamment lorsqu'elle est fondée sur des critères ethniques, génère des biais algorithmiques. De la même manière, si les données initiales sont déjà biaisées – parce qu'elles proviennent du Web où les discours racistes ou sexistes sont fréquents par exemple – les algorithmes produisent logiquement des résultats biaisés[18].

Parce qu'ils alimentent des discriminations, ces biais peuvent toutefois avoir de graves conséquences sur les individus et les communautés marginalisées[19]. Les exemples les plus saillants de ce phénomène sont les prédictions automatisées sur la propension à la récidive chez des prisonniers qui ont eu un impact sur les décisions de libération conditionnelle ; ou certains algorithmes utilisés pour l'aide au recrutement et qui ont systématiquement favorisé les hommes au détriment des femmes[20]. Une étude consacrée aux bases de données d'images exploitées par les forces de l'ordre aux États-Unis a quant à elle révélé que les individus d'ascendance africaine étaient plus susceptibles d'être incriminés à tort par les systèmes de reconnaissance faciale[21]. Ce biais s'explique par une surreprésentation des personnes d'ascendance africaine dans les bases de données photographiques de la police.

Une autre forme fréquente de biais dans les outils d'IA provient de la manière dont ceux-ci sont conçus. Même si les données utilisées pour alimenter un algorithme sont parfaitement représentatives, les choix de conception peuvent entraîner des résultats déformés et générer des effets discriminatoires importants. Par exemple, dans le cadre de l'élaboration d'un algorithme destiné à évaluer le risque de crédit, la définition et la mesure des facteurs de vulnérabilité peuvent aboutir à des résultats

16. A/HRC/56/68, « Rapport de la Rapporteuse spéciale sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée », Conseil des droits de l'Homme, Assemblée générale des Nations unies, 3 juin 2024.
17. E. Ferrara, « The Butterfly Effect in Artificial Intelligence Systems: Implications for AI Bias and Fairness », *Machine Learning With Applications*, vol. 15, 2024.
18. Ainsi, lorsqu'en 2016 avait été lancé Tay, le premier agent conversationnel de Microsoft, l'entreprise américaine avait été contrainte de mettre un terme à ses interactions sur les réseaux sociaux pour endiguer ses sorties racistes et sexistes.
19. N. Mehrabi *et al.*, « A Survey on Bias and Fairness in Machine Learning », *ACM Computing Surveys*, vol. 54, n° 6, 2022.
20. J. Angwin *et al.*, « Machine Bias », *ProPublica,* 13 mai 2016.
21. Citée dans C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Londres, Penguin Books, 2017.

biaisés[22]. Ainsi, le fait d'utiliser les cotes de crédit comme indicateur principal peut désavantager des groupes de personnes ayant généralement des cotes de crédit plus faibles.

Plus généralement, l'automatisation de certaines arbitrages – concernant l'accès à l'emploi, aux soins de santé, aux prêts bancaires ou aux programmes d'éducation – suscite des questionnements éthiques sur la transparence et la responsabilité des systèmes. L'un des défis majeurs de l'IA réside en effet dans la prise de décisions sans intervention humaine, souvent perçue comme se déroulant dans une « boîte noire ». Certains programmes fondés sur l'IA peuvent par exemple effectuer des choix de manière autonome, en se mettant à jour au fur et à mesure de leur exposition à de nouvelles données. Or, ces mises à jour peuvent progressivement les conduire à s'appuyer sur des critères différents de ceux initialement programmés, en fonction des tendances identifiées dans les données.

À mesure que ces nouvelles tendances influencent les décisions de l'algorithme, il devient de plus en plus difficile pour les utilisateurs de comprendre les facteurs sous-jacents aux résultats produits. Ce manque de transparence rend les processus de raisonnement des systèmes d'autant plus insaisissables et opaques[23]. Par ailleurs, de nombreux programmes développés par des entreprises échappent à tout examen juridique extérieur, notamment en raison des lois sur les contrats et la propriété intellectuelle. Cette absence de contrôle renforce les préoccupations liées au principe de responsabilité, et complique encore la régulation de ces technologies.

La question du respect de la vie privée constitue un deuxième enjeu essentiel[24]. Les systèmes d'IA s'appuient souvent sur des données contenant des informations personnelles, et leur collecte ou leur traitement sans le consentement des utilisateurs constitue une atteinte au droit à la vie privée[25]. De plus, il arrive que des données initialement recueillies dans un contexte spécifique, comme celui de la santé (par exemple dans le cas des applications médicales), soient partagées sans l'accord des individus concernés et utilisées dans d'autres domaines, tels que la justice. Par ailleurs, les fuites de données et les accès non autorisés, notamment par le

22. A/HRC/56/68, « Rapport de la Rapporteuse spéciale sur les formes contemporaines de racisme, de discrimination raciale, de xénophobie et de l'intolérance qui y est associée », *op. cit.*
23. J. Burrell, « How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms », *Big Data & Society*, vol. 3, n° 1, 2016
24. D'ailleurs, le cadre actuel de protection des données personnelles pourrait rapidement montrer ses limites en raison de la complexité de son application avec l'IA générative. Les alertes relatives au non-respect du Règlement général de protection des données (RGPD) se sont multipliées en Italie, en Espagne et en France, par exemple.
25. S. Lai et B. Tanner, « Examining the Intersection of Data Privacy and Civil Rights », Brookings Institution, 18 juillet 2022.

biais de piratages informatiques, suscitent de nouvelles inquiétudes quant à la protection des informations personnelles[26].

Ces atteintes à la vie privée sont d'autant plus problématiques que les outils d'IA peuvent également être utilisés dans des systèmes de contrôle social et de surveillance de masse[27]. Selon l'AI Global Surveillance Index, en 2019 déjà, 56 pays utilisaient l'IA dans leurs systèmes de surveillance des villes. Les États autocratiques peuvent alors utiliser la surveillance fondée sur l'IA pour détecter et suivre les individus dans le but de dissuader la désobéissance civile avant qu'elle ne commence, renforçant ainsi leur autorité. En Chine, par exemple, le gouvernement emploie ce genre de systèmes pour surveiller les minorités ethniques, notamment les Ouïghours, dans le cadre de son Sharp Eyes Program[28]. Dans un autre registre, l'entreprise américaine Clearview AI a récupéré des milliards de photographies d'individus sur les réseaux sociaux sans autorisation, ce qui lui a valu d'être poursuivie aux États-Unis et en Europe[29]. Ces pratiques mettent en effet gravement en péril la protection des données personnelles et des droits fondamentaux des citoyens.

Troisièmement, les effets combinés de l'IA, de l'automatisation et de la robotique dans de nombreux champs d'activité transforment rapidement le paysage du travail[30]. Un rapport du cabinet de conseil en stratégie McKinsey de 2024 prévoyait ainsi que jusqu'à 45 % des heures travaillées pourraient être automatisées d'ici 2035, notamment dans des secteurs comme la logistique, la comptabilité et la santé[31]. La même année, la banque d'investissement Goldman Sachs estimait pour sa part que 300 millions d'emplois à temps plein seraient menacés dans le monde[32]. Or, ces destructions d'emplois sont de nature à accroître les inégalités sociales, dans la mesure où les travailleurs peu qualifiés – qui exercent les métiers les plus répétitifs et donc automatisables – sont les premiers à en pâtir[33].

26. J. King et C. Meinhardt, « Rethinking Privacy in the AI Era: Policy Provocations for a Data Centric World », *White Paper*, HAI, Stanford University, février 2024.
27. A. Olvera, « How AI Surveillance Threatens Democracy Everywhere », *Bulletin of the Atomic Scientists*, 7 juin 2024.
28. D. Peterson, « How China Harnesses Data Fusion to Make Sense of Surveillance Data », Brookings Institution, 23 septembre 2021.
29. F. Filloux, « Reconnaissance faciale : Clearview AI, le poison mortel de la vie privée », *L'Express*, 18 octobre 2023.
30. D. Acemoglu et P. Restrepo, « The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand », *Cambridge Journal of Regions, Economy and Society*, vol. 13, n° 1, 2020 ; J. Nocetti, « L'Europe face à la numérisation du travail : quels risques politiques ? », *Études de l'Ifri*, Ifri, septembre 2018.
31. D. Barroux, « Presque la moitié des heures travaillées peuvent être automatisées par l'IA », *Les Échos*, 24 mai 2024.
32. N. Beyler, « ChatGPT et l'IA menacent 300 millions d'emplois dans le monde, selon Goldman Sachs », *Les Échos*, 28 mars 2023.
33. Les « cols blancs » dont les emplois sont également affectés peuvent plus facilement se réorienter pour intégrer les transformations technologiques.

Le développement rapide de ces technologies risque par conséquent d'aggraver des disparités déjà profondes, avec des répercussions sociopolitiques significatives. Ces répercussions sur l'emploi seraient d'autant plus graves que les sociétés peinent à mettre en place des solutions durables pour répondre aux besoins des populations les plus vulnérables ou marginalisées[34], qui se verraient encore davantage déclassées. Certains magnats de la Silicon Valley, à l'image de Sam Altman ou avant lui Bill Gates et Mark Zuckerberg, appellent de leurs vœux la mise en place d'un revenu universel de base, c'est-à-dire une allocation minimum qui compenserait la perte des emplois afin de pallier les troubles sociaux potentiels d'une paupérisation brutale d'une partie de la population[35].

Quatrièmement, d'importantes questions se posent concernant la propriété intellectuelle, l'IA générative faisant voler en éclat l'acception traditionnelle du droit d'auteur[36]. En effet, les LLM sont entraînés à partir de données récoltées en ligne, souvent sans consentement ni compensation de leurs créateurs. L'enjeu est alors de déterminer si l'utilisation à grande échelle de données pour entraîner des LLM peut être considérée comme relevant du « *fair use* », comme le prétendent les grands acteurs de l'IA générative, en particulier OpenAI[37]. Ce principe du droit d'auteur anglo-saxon autorise certaines utilisations sans consentement préalable de l'auteur, à condition que l'œuvre soit substantiellement transformée. Cette question s'applique donc à la génération automatique de contenus reproduisant le style d'artistes spécifiques, qui constituent une infraction au droit d'auteur.

C'est la raison pour laquelle, depuis 2023, de nombreux artistes, auteurs, humoristes, développeurs, maisons de disques et groupes médiatiques ont intenté des poursuites judiciaires contre différents géants du numérique américains – en particulier OpenAI, Microsoft, Stability AI, Midjourney, Meta et Anthropic pour ne pas les citer. En décembre 2023, le *New York Times* a quant à lui attaqué OpenAI et Microsoft en justice, pour avoir exploité le contenu de ses articles sans son consentement, afin d'entraîner ChatGPT et Copilot[38]. Il leur reproche également d'avoir reproduit intégralement – sans transformation – plusieurs de ses articles, citant des exemples précis. Ces pratiques, selon le journal, dépassent les limites généralement admises par la doctrine du *fair use*. Le quotidien soutient également que cela porte préjudice à ses relations avec ses lecteurs et impacte négativement ses principales sources de revenus, notamment les abonnements, les publicités et les partenariats.

34. K. Georgieva, « AI Will Transform the Global Economy: Let's Make Sure It Benefits Humanity », *IMF Blog*, 14 janvier 2024.
35. OpenAI estime ainsi que par moins de 80 % des travailleurs aux États-Unis pourraient voir leur emploi affecté par l'IA. Lire S. Emerson, « Sam Altman et le revenu universel de base », *Forbes*, 20 juillet 2024.
36. C. Metz, « Lawsuit Takes Aim at the Way A.I. Is Built », *The New York Times*, 23 novembre 2022.
37. « Does Generative Artificial Intelligence Infringe Copyright? », *The Economist*, 2 mars 2024.
38. « Le "New York Times" poursuit en justice Microsoft et OpenAI, créateur de ChatGPT, pour violation de droits d'auteur », *Le Monde*, 27 décembre 2023.

Ces affaires soulignent la nécessité d'un cadre juridique clair pour protéger les créateurs tout en favorisant l'innovation.

Enfin, l'IA engendre un coût environnemental colossal. Les centres de données utilisés pour entraîner et opérer les LLM consomment d'énormes quantités d'énergie. La popularité de ces modèles entraîne une consommation significative de ressources, de leur développement à leur utilisation. La fabrication de puces, le stockage des données, l'entraînement des modèles, les requêtes des utilisateurs et les données produites ont des effets notables sur les ressources physiques, hydriques et énergétiques, avec des répercussions directes sur le climat.

Selon une étude de l'université du Massachusetts, l'entraînement nécessaire à la mise au point d'un seul de ces LLM peut émettre autant de dioxyde de carbone (CO_2) que cinq voitures tout au long de leurs vies[39]. Par ailleurs, les besoins en eau pour refroidir les *data centers* demeurent opaques, mais il est probable qu'ils soient également considérables[40]. La mise en œuvre d'une norme mondiale harmonisée en matière de transparence, supervisée par les gouvernements nationaux, rendrait accessibles les données environnementales aux chercheurs et aux journalistes. Une telle approche offrirait au public la possibilité de scruter attentivement la consommation de ressources naturelles par les entreprises d'IA et permettrait aux décideurs politiques de mettre en place des restrictions pertinentes et efficaces. La gouvernance de l'IA ne saurait donc être dissociée des efforts de protection de l'environnement.

Pour la paix et la stabilité internationale

Un second volet de préoccupation majeure tient aux menaces que l'IA ferait peser sur la paix et la stabilité internationale. À cet égard, les risques en matière de cybersécurité sont particulièrement prégnants. D'une part, les acteurs cyberoffensifs tirent d'ores et déjà profit des opportunités offertes par les LLM en matière de production de code informatique, de traduction automatique, ou de curation d'éléments techniques. Ces procédés, déjà observés chez des groupes chinois, iraniens, nord-coréens ou russes[41], leur permettent de démultiplier leurs ressources en termes d'ingénierie sociale, d'usurpation d'identité et de manipulation grâce à un meilleur ciblage et des pratiques d'hameçonnages plus sophistiquées[42]. D'autre part, les

39. K. Haro, « Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes », *MIT Technology Review*, 6 juin 2019.
40. Début 2024, ils ont d'ailleurs été mis en cause au cours des mégafeux de Los Angeles, en raison de l'épuisement des ressources en eau. Lire T. Katzenberger, « AI Data Centers Face Scrutiny for Water and Energy Use as LA Fires Rage », *Politico*, 9 janvier 2025.
41. « Staying Ahead of Threat Actors in the Age of AI », Microsoft, 14 février 2024, disponible sur : www.micrsoft.com.
42. J. Hazell, « Spear Phishing with Large Language Models », Oxford Internet Institute, 14 décembre 2023, disponible sur : www.governance.ai.

technologies d'IA reposant sur des systèmes informatiques, elles constituent nécessairement une voie d'accès électromagnétique par laquelle des données peuvent être transmises pour mener une opération hostile. Leur généralisation risque donc d'étendre la surface d'attaque, à la fois des systèmes et des réseaux dans lesquels ils évoluent[43].

Si l'on en croit le National Cyber Security Centre (NCSC) britannique, les LLM s'avéreraient notamment vulnérables à deux catégories d'attaques :

- les instructions malveillantes (*prompt injections*) dont l'objectif est de manipuler le modèle par le biais de son interface de commande ;
- l'empoisonnement des jeux de données (*data poisoning*), qui peut advenir en amont ou au cours de l'entraînement des modèles.

Parce que ceux-ci sont élaborés à partir d'ensembles massifs de données ouvertes et qu'ils tendent à être exploités pour fournir des informations à des applications et services tiers, les attaques visant à corrompre ces jeux de données représentent un risque significatif[44].

Ces capacités nouvelles font craindre la prolifération de cyberattaques de type rançongiciel ou par déni de service distribué (DDoS), dans la mesure où les techniques d'IA permettent également de mieux coordonner les attaques grâce à des réseaux de systèmes contaminés (*botnets*). L'industrie du crime organisé s'est déjà pleinement saisie de ces outils émergents[45] et les LLM sont désormais légions sur le *dark web*[46]. Il y a donc fort à parier que les puissances étatiques particulièrement investies dans la lutte informatique offensive ne manqueront pas d'en tirer profit.

Deuxièmement, et associées au risque cyber, se font jour des interrogations sur l'intégrité de l'information à l'ère des LLM. Au-delà des seules « hallucinations » observées chez les *chatbots*, qui tendent parfois à inventer ou déformer des renseignements[47], l'IA générative suscite de nombreuses craintes en matière de désinformation. Celle-ci permet en effet d'amplifier considérablement les capacités de manipulation de l'information, à la fois en termes de production et de diffusion, notamment dans le cadre de campagnes d'ingérences numériques étrangères. Les contenus informationnels fabriqués de toutes pièces à l'aide de l'IA générative, appelés *deepfakes*, prolifèrent ainsi sur les réseaux sociaux, avec des conséquences concrètes dans le monde réel.

43. J. Jun, « How Will AI Change Cyber Operations », War on the Rocks, 30 avril 2024.
44. B. Pajot, « Les risques de l'IA : enjeux discursifs d'une technologie stratégique », *op. cit.*
45. D. Larousserie, « Comment les *chatbots* ont été gangrenés par l'industrie du crime organisé », *Le Monde*, 13 février 2024.
46. J. Cheminat, « Après WormGPT, les cybercriminels livrent FraudGPT », *Le Monde Informatique*, 26 juillet 2023.
47. C. Metz, « Chatbots May "Hallucinate" More Often Than Many Realize », *The New York Times*, 6 novembre 2023.

Le 19 octobre 2023, le journal *Libération* affichait ainsi en « une » la photographie d'un manifestant au Caire brandissant l'image d'un bébé en larmes dans les décombres de l'hôpital gazaoui Al-Ahli Arabi. Cette dernière avait en réalité été générée par IA et diffusée à l'occasion des tremblements de terre survenus à proximité de la frontière turco-syrienne en février 2023[48]. Bien qu'en l'espèce les photographies d'enfants dans les décombres de Gaza ne manquent pas, l'usage de cette image artificielle est venu renforcer les soupçons de *fake news* et de mise en scène des victimes palestiniennes – également appelée « Pallywood[49] ».

L'IA générative favorise aussi l'apparition de nouveaux procédés de lutte informatique d'influence « complexe ». En octobre 2023 aux Émirats arabes unis, des hackers pro-iraniens ont ainsi interrompu le journal télévisé pour y diffuser un faux reportage généré par l'IA au sujet de la guerre à Gaza[50]. Dans un autre registre, l'élection présidentielle roumaine de décembre 2024, qui avait porté en tête, à l'issue du premier tour, le candidat ultra-nationaliste et complotiste Călin Georgescu, a dû être annulée en raison de soupçons d'instrumentalisation de TikTok par la Russie en sa faveur – le candidat étant favorable à l'arrêt immédiat de l'aide à l'Ukraine. Les modes opératoires de cette campagne d'ingérence informationnelle demeurent flous à ce stade, mais il est probable que l'IA générative ait servi à élaborer et amplifier des contenus pro-Georgescu sur la plateforme chinoise[51]. L'IA contribue ainsi à amplifier les tensions internationales, en servant les intérêts stratégiques des États désireux d'investir les sphères informationnelles de leurs adversaires afin d'y exercer une influence et de manipuler les perceptions.

Les techniques d'IA rendent donc les attaques informationnelles plus massives, plus sophistiquées et plus ciblées, ce qui leur permet d'atteindre une audience cible beaucoup plus importante tout en contournant les mécanismes de détection mis en place par les plateformes, par le biais de relais indirects, tels que des leaders d'opinion et des influenceurs qui partagent par inadvertance des contenus inauthentiques. De plus, bien qu'ils aient par le passé mis en scène leur détermination à lutter contre ces tentatives de désinformation[52], les *Big Tech* sont régulièrement critiqués

48. « "Libé" s'est-il rendu coupable d'une "fake news" en publiant la vraie photo d'un homme brandissant une image générée par IA ? », *Libération*, 19 octobre 2023.
49. W. Audureau, « "Pallywood" : en plein carnage à Gaza, le mythe des fausses morts palestiniennes », *Le Monde*, 16 décembre 2023.
50. D. Milmo, « Iran-backed Hackers Interrupt UAE TV Streaming Services with Deepfake News », *The Guardian*, 8 février 2024.
51. M. Bran, « En raison de l'influence de TikTok, les juges roumains annulent la présidentielle », *Le Monde*, 7 décembre 2024.
52. En février 2024, Google, Meta, Microsoft, OpenAI, TikTok et Adobe ont par exemple signé un accord de lutte contre les *deepfakes* en contexte électoral. Lire G. De Vynck, « AI Companies Agree to Limit Election "Deepfakes" But Fall Short of Ban », *The Washington Post,* 13 février 2024.

pour leur mauvaise gestion des contenus falsifiés[53]. Ces exemples répétés de manipulations de l'information mettent en exergue la nécessité de labelliser les contenus générés par IA, qui bien qu'elle soit désormais exigée par le *Digital Services Act* européen[54], en est encore à ses balbutiements[55].

Troisièmement, sur le champ de bataille aussi les promesses de l'IA font l'objet d'attentes grandissantes et d'innovations fulgurantes, ce qui n'est pas sans poser de nombreuses questions en matière de stratégie, de politique, de droit et d'éthique.

« *The First AI War* » : c'est ainsi que le magazine américain *Time* présentait la guerre russo-ukrainienne en couverture de son numéro du 26 février 2024. Les conflits en Ukraine et à Gaza ont en effet consacré l'IA comme un véritable multiplicateur de force : ses applications militaires sont multiples et protéiformes, de la logistique au ciblage en passant par le renseignement et l'aide à la décision au sein des fonctions de commandement et de contrôle (C2)[56]. Elle est même envisagée comme une nouvelle révolution des techniques de la guerre, au même titre qu'avant elle la poudre à canon ou l'arme nucléaire.

Les progrès de l'IA militaire couplés aux avancées en matière de robotique font toutefois redouter leur utilisation abusive à des fins militaires, en particulier dans le cas des systèmes d'armes létales autonomes (SALA), que les médias appellent volontiers « robots tueurs ». Ceux-ci désignent des systèmes qui, une fois activés, peuvent identifier une cible et user de la force létale sans être supervisés par un opérateur humain.

Se pose alors la question de la métacognition, si des systèmes poursuivaient leur apprentissage en cours de mission afin de s'adapter à des environnements changeants. Sans supervision efficace, ce qui serait « appris » en conduite pourrait donner lieu à des réactions inattendues, indésirables et ne correspondant pas au cadre d'emploi envisagé[57].

53. C. Zakrzewski, « ChatGPT Breaks Its Own Rules on Political Messages », *The Washington Post,* 28 août 2024.

54. En vertu de l'article 35-k, et si des risques systémiques sont recensés sur la plateforme, dont la manipulation des opinions, et les effets négatifs sur les processus démocratiques et électoraux, ainsi que les discours civiques.

55. K. Lentschner, « Les plateformes tâtonnent face à la labellisation des contenus IA », *Le Figaro*, 2 juillet 2024.

56. D'après l'hebdomadaire *Dzerkalo Tyjnia*, l'utilisation de l'IA par l'armée ukrainienne se décline désormais en dix domaines différents : l'autonomie des systèmes d'armes ; l'observation et la reconnaissance ; l'identification et la classification des cibles ; l'analyse et la prédiction des menaces ; la logistique et le ravitaillement ; la cybersécurité ; la guerre électronique ; la simulation et la formation ; la santé des armées ; et l'aide à la décision. Lire à ce sujet A. Férey et L. de Roucy-Rochegonde, « De l'Ukraine à Gaza : l'Intelligence artificielle en guerre », *Politique étrangère*, vol. 89, n° 3, Ifri, septembre 2024.

57. Par exemple, lors d'une simulation organisée par l'armée de l'Air américaine, un drone piloté par un programme d'IA aurait décidé de « tuer » son opérateur pour l'empêcher d'interférer dans ses efforts pour accomplir sa mission. La porte-parole de l'*US Air Force* a toutefois contesté l'existence de cette simulation. Lire « US Air Force Denies Running Simulation in Which AI Drone 'Killed' Operator », *The Guardian*, 2 juin 2023.

Plus largement, des systèmes auto-apprenants et capables d'évoluer au cours de leur utilisation interrogent, au-delà même de la maîtrise de leur configuration, sur la possibilité d'en garantir la fiabilité dans la durée.

La Stratégie française d'intelligence artificielle de défense avance par ailleurs que des puissances révisionnistes, telles que la Russie et la Chine, misent sur l'IA militaire pour bousculer le *statu quo* international à leur avantage. Les innovations dans ce domaine peuvent en effet favoriser un nivellement des positions stratégiques, car elles sont relativement peu coûteuses et aisées à maîtriser. Dans le même temps, d'autres acteurs pourraient « rentrer dans le jeu » en acquérant des technologies qui, bien que complexes, deviennent de moins en moins onéreuses et donc toujours plus accessibles. Cette dissémination permettrait alors aux parties les plus faibles de modifier les rapports de force internationaux.

L'arsenalisation de l'IA suscite donc de nombreuses inquiétudes. En novembre 2018, à l'occasion du premier Forum de Paris sur la Paix, le Secrétaire général de l'Organisation des Nations unies (ONU), Antonio Guterres, avait ainsi appelé à ce que soient « interdites par la législation internationale ces armes politiquement inacceptables et moralement révoltantes[58] ». En octobre 2023, avec la présidente du Comité international de la Croix-Rouge Mirjana Spoljaric Egger, il réitérait son plaidoyer pour leur interdiction. Les craintes portent notamment sur la compatibilité de tels systèmes avec le droit des conflits armés et le principe de dignité humaine, sur le risque d'abaissement du seuil d'entrée en conflit et d'escalades destructrices, et sur leur diffusion auprès d'acteurs non étatiques violents[59].

L'IA entraîne de surcroît un risque d'abaissement du seuil technologique d'accès aux armes de destruction massive. Des modèles d'IA pourraient en effet être détournés pour créer des formules chimiques ou biologiques inédites et potentiellement plus dangereuses que toutes celles connues, au profit d'utilisateurs non autorisés (groupes terroristes, criminels…). Des acteurs étatiques malveillants pourraient également être en mesure de développer de nouvelles armes, renforcer leur létalité et rendre plus difficile l'identification de l'agent employé, freinant logiquement le développement et l'administration de contre-mesures adaptées.

C'est la raison pour laquelle la régulation de l'IA militaire fait désormais l'objet de discussions dans les enceintes de négociation multilatérales. À l'échelle internationale, le débat sur les armes autonomes a été engagé par des ONG coalisés au sein d'une *Campaign to Stop Killer Robots* en 2012. Il a ensuite été discuté au Conseil des droits de l'Homme, puis dans le cadre de la Convention sur certaines armes classiques, où en

58. « Allocution du Secrétaire général au Forum de Paris sur la Paix », Paris, 11 novembre 2018.
59. L. de Roucy-Rochegonde, *La Guerre à l'ère de l'Intelligence artificielle : quand les machines prennent les armes*, Paris, PUF, 2024.

2016 a été décidée la création d'un groupe d'experts gouvernementaux doté d'un mandat de discussion sur la question, renouvelé en 2023. Plus récemment, lors de leur dernière rencontre officielle en novembre 2024, les présidents Xi et Biden se sont accordés sur la nécessité de limiter l'intégration de l'IA aux systèmes d'armes nucléaires[60].

Il s'agit donc d'établir des limites claires concernant l'utilisation d'armes reposant sur l'IA, y compris les cyberarmes. La mise en œuvre du droit international dans le contexte des cyberopérations est déjà un domaine mal défini[61], auquel l'IA ajoute encore de la complexité. Cette technologie renforce les avantages des acteurs cyberoffensifs en leur permettant, par exemple, d'utiliser l'IA générative pour analyser rapidement de vastes volumes de logiciels et identifier leurs failles. Un accord international interdisant certaines utilisations de l'IA militarisée, telles que les armes autonomes ou la propagation de désinformation pendant les campagnes électorales d'un autre pays, pourrait ainsi instaurer des garde-fous essentiels et promouvoir des pratiques responsables.

Répartir les externalités

Au-delà de la seule question des risques de l'IA se pose celle de la répartition des externalités qui y sont associées. D'une part, certains dénoncent la concentration du pouvoir au sein des grandes entreprises technologiques prévalant dans ce nouvel « été » de l'IA. D'autre part, son développement s'inscrit dans des processus transnationaux et engendre des externalités transfrontalières, ce qui rend nécessaire une coopération globale afin d'élaborer des cadres de régulation dépassant les frontières.

D'abord, les *foundation models* capables d'accomplir des tâches très diverses, qui sont au fondement des modèles secondaires et des applications d'IA connus aujourd'hui, sont élaborés par quelques géants technologiques. En effet, les coûts d'entrée exorbitants de ces modèles tendent à favoriser les primo-entrants, ce qui donne lieu à une forte concentration du marché[62]. Celle-ci est encore renforcée par les logiques de prédation des *Big Tech*, dont témoignent les investissements de Google dans Deepmind, de Microsoft dans OpenAI et Mistral AI ou d'Amazon dans Anthropic[63].

60. L. Egan et P. Kine, « Biden's Final Meeting with Xi Jinping Reaps Agreement on AI and Nukes », *Politico*, 16 novembre 2024.
61. F. Delerue, *Cyber Operations and International Law*, Cambridge, Cambridge University Press, 2020.
62. J. Vipra et A. Korinek, « Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT », Brookings Institution, 7 septembre 2023.
63. E. Ludlow, M. Day et D. Bass, « Amazon to Invest Up to $4 Billion in AI Startup Anthropic », *Bloomberg*, 25 septembre 2023.

Ce sont en effet ces mastodontes qui disposent des ressources stratégiques – fonds, semi-conducteurs, puissance de calcul, données, algorithmes, *cloud* et talents – qui leur donnent les leviers nécessaires pour infléchir les grandes orientations en matière d'IA[64]. Ces moyens colossaux leur permettent en outre d'attirer les profils les plus prometteurs dans le monde entier, dépouillant au passage la recherche académique et les institutions publiques[65], d'autant qu'ils n'hésitent pas à débaucher les meilleurs éléments chez des concurrents plus modestes, à l'image de Stability AI[66], et que leurs pratiques anticoncurrentielles et monopolistiques ne sont plus à démontrer[67]. Comme le résume le président de l'Autorité de la concurrence française, Benoît Cœuré : « L'IA est la première technologie à être d'emblée dominée par des grands acteurs[68]. » C'est en effet la première innovation de rupture dans laquelle les entreprises les plus puissantes contrôlent à elles seules les capacités de la développer. C'est la raison pour laquelle de plus en plus d'acteurs appellent de leurs vœux une politique combinant régulation et investissements publics dans l'IA, afin de contrebalancer l'influence grandissante des acteurs privés[69].

Parfois présentée comme une alternative à cette hégémonie des *Big Tech*, l'*open source* n'est pas non plus exempt de risques. Certes, les innovations « ouvertes » sont moins exposées à la pression commerciale, qui pousse les acteurs propriétaires à avancer le plus rapidement possible afin de conserver leur avantage concurrentiel, au risque que leurs modèles ne soient ni totalement aboutis ni suffisamment sécurisés. Cependant, précisément du fait de leur ouverture, elles offrent des outils extrêmement puissants à n'importe quel acteur, y compris malveillant, ce qui n'est pas sans susciter des inquiétudes[70]. En novembre 2024, il a ainsi été révélé que la Chine avait eu recours au modèle Llama 13B de Meta pour mettre au point un *chatbot* à vocation militaire[71]. Les modèles ouverts sont également plus exposés aux risques cyber, notamment en matière d'intoxication et d'empoisonnement des données.

64. A. Kak, S. Myers West et M. Wittaker, « Make No Mistake, AI Is Owned by Big Tech », *MIT Technology Review,* 5 décembre 2023.
65. N. Nix, C. Zakrzewski et G. De Vynck, « Silicon Valley Is Pricing Academics Out of AI Research », *The Washington Post*, 10 mars 2024.
66. T. Warren, « Stability AI CEO Resigns to "Pursue Decentralized AI" », *The Verge*, 23 mars 2024.
67. D. Milmo, « UK Watchdog to Examine Microsoft's Partnership with OpenAI », *The Guardian,* 8 décembre 2023.
68. A. Piquard, « L'IA est la première technologie à être d'emblée dominée par les grands acteurs », *Le Monde*, 27 septembre 2024.
69. M. Schaake, « AI Is Too Important to Be Monopolised », *Financial Times*, 12 décembre 2024.
70. Celles-ci sont notamment portées par les acteurs politiques et économiques américains, car elles présentent le double avantage de mettre en garde contre la menace chinoise tout en renforçant l'assise des *Big Tech* propriétaires de modèles fermés.
71. J. Pomfret et J. Pang, « Chinese Researchers Develop AI Model for Military Use on Back of Meta's Llama", Reuters, 1er novembre 2024.

Par ailleurs, la répartition des coûts et des bénéfices de l'IA à l'échelle mondiale fait l'objet d'un clivage Nord/Sud. Les pays en développement et émergents s'avèrent en effet particulièrement vulnérables aux bouleversements sociaux induits par l'IA, du fait de la pression exercée sur les emplois peu qualifiés et à faible valeur ajoutée. Ces États ne bénéficient pas non plus d'une couverture sociale suffisante, ni de recours efficaces pour opérer un bond technologique et véritablement tirer profit du développement de l'IA[72].

Ces pays concentrent aussi une grande partie des « travailleurs du clic » indispensables à l'entraînement des modèles d'IA[73]. Or, leurs conditions de vie extrêmement précaires contrastent fortement avec les profits colossaux générés par cette industrie[74]. L'asymétrie dans la collecte de données, combinée à une répartition inéquitable des coûts et des bénéfices – l'IA étant soutenue par les ressources minières et la main-d'œuvre issues des pays du Sud, mais conçue principalement par et pour les populations des pays du Nord – nourrit un profond sentiment d'injustice. Ce déséquilibre est renforcé par le manque criant de représentativité des modèles d'IA, souvent entraînés sur des données provenant du monde occidental, majoritairement produites par des hommes à hauts revenus, et rédigées en anglais. Ces biais systématiques véhiculent des stéréotypes susceptibles non seulement d'accentuer les inégalités sociales au niveau domestique, mais aussi d'aggraver les divergences politiques et culturelles à l'échelle internationale[75].

L'Organisation mondiale de la santé (OMS) a ainsi alerté sur les effets potentiellement néfastes du recours aux technologies d'IA dans le domaine de la santé auprès des populations des pays en développement. Elle a notamment souligné le problème du manque de diversité dans les données d'entraînement, qui limite leur efficacité pour des groupes insuffisamment représentés. L'organisation a également exprimé son inquiétude quant à la domination du secteur privé, au détriment de la recherche académique et des initiatives des agences publiques.

Dans ce contexte, les appels à instaurer des mécanismes de régulation pour éviter d'exacerber les clivages se font de plus en plus pressants, y compris au niveau international[76]. Parmi les pistes explorées figure l'introduction d'une taxe dédiée pour compenser les impacts sociaux de

72. D. Björkegren, « Artificial Intelligence for the Poor: How to Harness the Power of AI in the Developing World », *Foreign Affairs*, 9 août 2023.
73. G. Kristanadjaja, « Intelligence artificielle : dans les pays du Sud, des petites mains victimes de "colonisation numérique" », *Libération*, 21 mars 2024.
74. B. Perrigo, « OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic », *Time*, 18 janvier 2023.
75. V. Türk, « How AI Reduces the World to Stereotypes », *Rest of World*, 10 octobre 2023.
76. L. Eliott, « Big Tech Firms Recklessly Pursuing Profits from AI, Says UN Head », *The Guardian*, 17 janvier 2024.

l'IA[77]. Au regard de l'asymétrie flagrante dans la répartition des externalités positives et négatives de l'IA[78], certains plaident pour une gouvernance « technoprudentielle » ou pour un *containment* de ces technologies[79]. L'objectif est de mobiliser des décideurs, des acteurs industriels et des représentants de la société civile pour garantir un contrôle collectif de ces outils tout en veillant à la répartition équitable de leurs externalités à l'échelle mondiale.

La gouvernance émergente de l'IA aura en effet des implications sur la façon dont les avantages et les charges sont répartis entre les groupes sociaux et les États[80]. À l'instar de la réglementation des innovations technologiques antérieures[81], la gouvernance de l'IA peut en ce sens produire des avantages collectifs, ou au contraire favoriser certains acteurs au détriment d'autres[82]. La nécessité d'une réponse globale découle alors du caractère universel de ces externalités, notamment dans les domaines de la science et de l'innovation. Les institutions multilatérales doivent faciliter le partage des connaissances, la mise en commun des ressources et la coordination des efforts de recherche, tout en veillant à ce que les avancées profitent équitablement à l'ensemble de la planète.

Maîtriser les conséquences sur d'autres enjeux globaux

Le troisième volet de préoccupations plaidant pour une gouvernance mondiale de l'IA tient aux conséquences que celle-ci emporte sur d'autres enjeux globaux. En effet, de nombreux gouvernements et organismes publics ont déjà intégré l'IA dans leurs activités quotidiennes, afin d'évaluer plus efficacement l'éligibilité à l'aide sociale, signaler les fraudes potentielles, établir le profil de suspects, évaluer les risques et exercer une surveillance[83].

77. M. Schaake, « It's Already Time to Think About an AI Tax », *Financial Times,* 8 janvier 2024, disponible sur : www.ft.com.
78. I. Bremmer et M. Suleyman, « The AI Power Paradox: Can States Learn to Govern Artificial Intelligence Before It's Too Late? », *Foreign Affairs,* 16 août 2023.
79. M. Suleyman, « Containment for AI: How to Adapt a Cold War Strategy to a New Threat », *Foreign Affairs*, 23 janvier 2024.
80. R. Gilpin, *The Political Economy of International Relations*, Princeton, Princeton University Press, 1987 ; A. Dreher et V. Lang, « The Political Economy of International Organizations », *in* R. Congleton, B. Grofman et S. Voigt (dir.), *The Oxford Handbook of Public Choice*, Oxford, Oxford University Press, 2019.
81. D. Drezner, « Technological Change and International Relations », *International Relations*, vol. 33, n° 2, 2019.
82. A. Dafoe, *AI Governance: A Research Agenda*, Governance of AI Program, Future of Humanity Institute, University of Oxford, 2018.
83. G. Misuraca et C. van Noordt, « Artificial Intelligence for the Public Sector: Results of Landscaping the Use of AI in Government Across the European Union », *Government Information Quarterly*, vol. 39, n° 3, 2022.

Les systèmes d'IA ne sont pourtant pas infaillibles : il est en réalité fréquent qu'ils commettent des erreurs, avec des effets de bords parfois dramatiques. Les autorités néerlandaises ont récemment mis en œuvre un algorithme qui a plongé des dizaines de milliers de familles dans la pauvreté après leur avoir demandé par erreur de rembourser les allocations familiales, ce qui a finalement contraint le gouverneur à démissionner[84]. En Australie, le système *Robodebt* conçu pour détecter les paiements de sécurité sociale erronés, a quant à lui émis à tort 400 000 dettes sociales que le gouvernement australien a dû annuler[85].

Il est estimé que pas moins de 85 % de tous les projets intégrant de l'IA provoqueront des erreurs dues aux biais des algorithmes, de leurs développeurs ou des données utilisées pour les faire fonctionner[86]. La base de données AI Incident Database recense d'ailleurs les cas d'incidents provoqués par ce type d'erreurs et les maux qu'ils ont causés[87]. La question se pose alors des réparations à apporter à ceux ayant subi un préjudice causé par un dysfonctionnement de l'IA.

Il convient par ailleurs de rappeler que les techniques d'IA et leurs applications variées – allant du marketing aux soins de santé en passant par les systèmes d'armes – sont appelées à transformer en profondeur la société et le monde dans son ensemble. Il s'avère donc d'autant plus difficile d'encadrer cette technologie polyvalente, dont les effets s'observent dans des champs si divers. De la même manière que l'adoption de l'IA peut entraîner des répercussions sur l'ensemble de l'économie mondiale, les efforts de réglementation doivent donc aller au-delà des questions liées à la seule technologie.

Dans le même temps, le développement planétaire de l'IA introduit de nouvelles sources de rivalités et de tensions. Ces technologies pourraient exacerber les inégalités économiques à la fois au sein des États et entre eux, posant des menaces à la sécurité internationale. Des applications mal régulées risquent de perturber la stabilité nucléaire, de faciliter l'élaboration d'armes biologiques et chimiques, ou encore de démocratiser l'usage de systèmes d'armes autonomes. Or, l'autre raison d'être de toute gouvernance mondiale est précisément d'empêcher l'hostilité entre les nations de déboucher sur des conflits ouverts.

Pour répondre à ces défis, il est impératif de réguler la manière dont les technologies d'IA sont conçues, diffusées et utilisées, afin de préserver les intérêts de tous. Comme pour d'autres technologies à usage général – telles

84. M. Heikkilä, « AI: Decoded: A Dutch Algorithm Scandal Serves a Warning to Europe — The AI Act Won't Save Us », *Politico*, 30 mars 2022.
85. L. Henriques-Gomes, « Robodebt: Government Admits It Will Be Forced to Refund $550m under Botched Scheme », *The Guardian*, 26 mars 2020.
86. « Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence », *Gartner Newsroom*, 13 février 2018, disponible sur : www.gartner.com.
87. Disponible sur : https://incidentdatabase.ai.

que la machine à vapeur, l'électricité ou encore Internet[88] – l'IA influence profondément la compétitivité économique, la sécurité militaire et l'intégrité individuelle, avec des conséquences pour les États et les sociétés[89].

Un large éventail de politiques cohérentes avec les décisions prises en matière d'encadrement de l'IA doit alors être mis en œuvre par les régimes de gouvernance existants et émergents. Le régime commercial international pourrait par exemple être repensé pour prendre en compte les évolutions liées à l'IA et ses spécificités[90]. De la même manière, le Fonds international pour la diversité culturelle (FIDC), qui dépend de l'Organisation des Nations unies pour l'éducation, la science et la culture (UNESCO), a probablement un rôle à jouer dans la défense de la diversité linguistique, largement mise à mal par la prépondérance de l'anglais dans les LLM. Plus généralement, les débats sur l'IA croisent des enjeux tels que l'équité du développement économique ou le respect des droits humains, qui font déjà l'objet de travaux et d'investissements par la communauté internationale.

Les défis auxquels sont confrontés la plupart des pays et qui touchent à des sphères aussi variées requièrent une réponse coordonnée à l'échelle mondiale. La gouvernance mondiale joue en outre un rôle fondamental dans le renforcement des échanges culturels et la compréhension mutuelle entre les nations. Des institutions comme l'UNESCO permettent ainsi de bâtir des ponts entre les cultures, de promouvoir le dialogue et de nourrir un sentiment d'appartenance à une communauté mondiale.

Une gouvernance efficace de l'IA ne saurait néanmoins se limiter à des cadres nationaux ou régionaux. Les gouvernements doivent collaborer pour instaurer des standards interopérables et coordonnés à l'échelle mondiale, fondés sur une compréhension rigoureuse des incidents et des dangers liés à l'IA. Cela inclut l'encadrement éthique de l'utilisation des données, la gestion des impacts environnementaux et la prévention des abus liés aux algorithmes. Une telle approche permettrait de garantir que l'IA profite à l'ensemble de la société tout en minimisant ses externalités négatives. Face à ces défis colossaux, les acteurs de l'IA et les régulateurs se sont mis en ordre de marche, mais avancent en ordre dispersé.

88. C. B. Frey, *The Technology Trap: Capital, Labor and Power in the Age of Automation*, Princeton, Princeton University Press, 2019.
89. J. Tallberg *et al.*, « The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research », *International Studies Review*, vol. 25, n° 3, 2023.
90. E. Erman et M. Furendal, « The Global Governance of Artificial Intelligence: Some Normative Concerns », *Moral Philosophy & Politics*, vol. 9, n° 2, 2022.

Une balkanisation de la gouvernance

Lors du sommet du G7 organisé du 13 au 15 juin 2024 dans la région des Pouilles en Italie, le pape François a exhorté les responsables politiques à « prendre des mesures concrètes pour gouverner le processus technologique en cours [dans le domaine de l'IA] dans le sens de la fraternité et de la paix[91] ». Si la nécessité d'élaborer des normes claires sur l'utilisation licite et éthique des données, la protection de la propriété intellectuelle ou la limitation des dommages environnementaux liés à l'IA semble à cet égard consensuelle, les applications concrètes supposées en découler demeurent floues.

Initiés en 2019[92], les efforts d'encadrement de l'IA ont pris de l'ampleur à partir du printemps 2021, lorsque la Commission européenne a présenté le plan initial de son *AI Act*. Parallèlement, la Chine et les États-Unis se sont eux aussi dotés de nouveaux dispositifs normatifs au sujet de l'IA, tandis que de nombreuses initiatives multilatérales ou venues de la société civile voyaient le jour.

Le paysage réglementaire actuel s'avère par conséquent éminemment fragmenté[93]. L'élaboration de nouvelles normes est discutée dans d'innombrables forums, alors même que les ressources diplomatiques sont limitées et qu'il est impossible pour un État de s'investir dans toutes les enceintes de négociation à la fois. Cette situation donne lieu à un phénomène de « *forum shopping* », qui voit les États choisir stratégiquement des enceintes avec plus ou moins de restrictions ou de surveillance, en fonction de leurs intérêts nationaux et des contraintes qu'ils veulent faire peser sur leurs compétiteurs[94]. En résulte une cacophonie normative, qui peine à s'harmoniser.

Cette deuxième partie expose donc la manière dont est actuellement régulée l'IA sur le plan international : où et comment de nouveaux dispositifs réglementaires émergent-ils au niveau mondial ?

---

91. L. Besmond de Senneville, « Au G7, le pape François en défenseur du "contrôle humain" face à l'IA », *La Croix*, 14 juin 2024.

92. Bien que de premières occurrences aient vu le jour du côté des acteurs privés dès 2017, avec notamment la création du AI for Good Global Summit, issu de d'un partenariat entre la fondation canadienne X Prize et l'Union internationale des télécommunications.

93. L. Schmitt, « Mapping Global AI Governance: A Nascent Regime in a Fragmented Landscape », *AI and Ethics*, vol. 2, 2022.

94. W. Hofmann-Riem, « Artificial Intelligence as a Challenge for Law and Regulation », *in* T. Wischmeyer et T. Rademacher (dir.), *Regulating Artificial Intelligence*, Cham, Springer, 2020.

Les approches des trois « blocs » de l'IA

La course à l'IA à l'échelle internationale s'articule autour de trois « blocs » principaux : les États-Unis, la Chine et l'Europe[95]. Or, dans cette compétition mondiale, la gouvernance est un enjeu tout aussi essentiel que ne le sont les investissements, les corpus de données, les algorithmes, la puissance de calcul ou encore les talents. Il s'agit en effet de se doter d'un arsenal normatif pour soutenir l'innovation.

Dans ce domaine, l'UE, reconnue comme l'une des principales puissances normatives, parvient à tirer son épingle du jeu, même si la scène mondiale en matière d'innovation reste dominée par le duopole sino-américain. En quoi les approches de ces trois « blocs » diffèrent-elles alors ?

*L'*AI Act *européen : vers un nouvel « effet Bruxelles » ?*

L'UE s'est révélée pionnière en matière de régulation de l'IA. Dès 2020 la Commission publiait son *Livre blanc sur l'intelligence artificielle*, donnant lieu à des débats puis à une proposition officielle de législation européenne en la matière le 21 avril 2021. Le 6 décembre 2022 le Conseil européen en adoptait une orientation générale puis entamait des négociations avec le Parlement, avant de parvenir à un accord un an plus tard, le 3 décembre 2023. Le 13 mars 2024, le projet de règlement était ainsi adopté par la neuvième législature du Parlement européen, par 523 voix pour et 46 contre. Parallèlement, le 17 mai 2024 était par ailleurs adoptée la Convention-cadre du Conseil de l'Europe sur l'intelligence artificielle et les droits de l'homme, la démocratie et l'État de droit[96]. Puis, le 21 mai 2024, le texte était officiellement adopté par les 27 ministres réunis en Conseil de l'UE. Publié dans les semaines suivantes au *Journal Officiel* de l'UE, l'*AI Act* est entré en vigueur le 1er août 2024 et doit être progressivement mis en application par l'*Office AI* – spécialement créé dans ce but en janvier 2024 – entre février 2025 et août 2027[97]. Le 14 novembre 2024, la Commission a publié le premier projet de code de bonnes pratiques en matière d'IA à

95. B. Pajot, « Intelligence artificielle : la compétition internationale », *op. cit.*
96. Ce qui en fait le premier instrument international juridiquement contraignant, ouvert à la signature le 5 septembre 2024. L'ont jusqu'à présent signé Andorre, la Géorgie, l'Islande, le Monténégro, la Norvège, la Moldavie, le Royaume-Uni, Saint-Marin, les États-Unis, Israël et l'Union européenne.
97. En février 2025, le chapitre I correspondant aux dispositions générales et le chapitre II correspondant à la pratique interdite en matière d'IA, c'est-à-dire les applications d'IA à risque « inacceptable » s'appliqueront. En août 2025, le chapitre III section 4 (Autorités notifiantes et organismes notifié), le chapitre V (modèles d'IA à usage général), le chapitre VII (gouvernance), le chapitre XII (sanctions) contenant l'article 78 (confidentialité) s'appliqueront, à l'exception de l'article 101 (Amendes applicables aux fournisseurs). En août 2026, l'ensemble du règlement s'appliquera, à l'exception de l'article 6, paragraphe 1 du chapitre III et des obligations correspondant aux catégories de systèmes d'IA à risque « élevé ». En août 2027, tout le règlement doit s'appliquer.

finalité générale, qui vise à faciliter la bonne mise en œuvre des règles définies par l'*AI Act*[98].

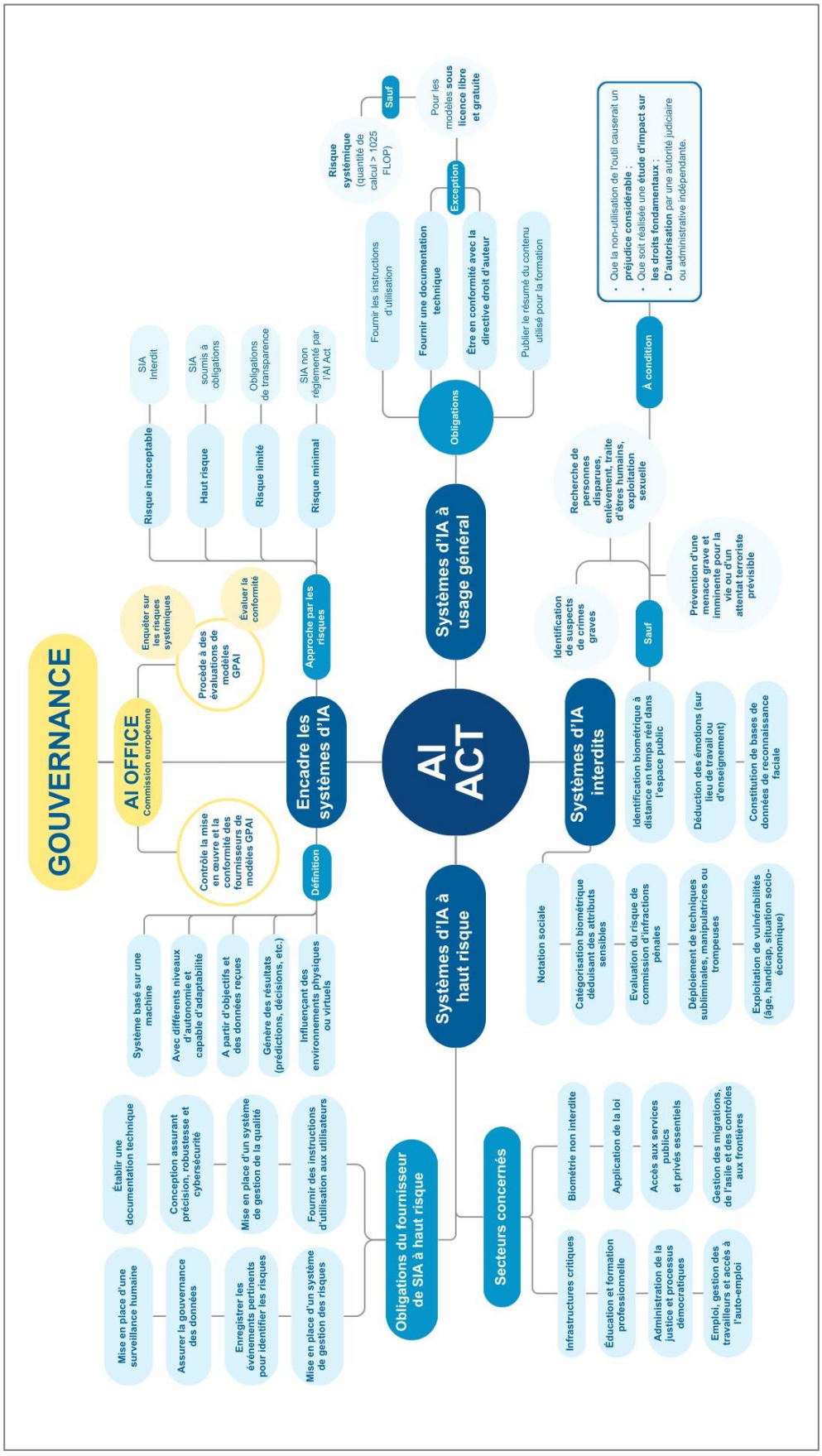
Premier cadre réglementaire contraignant sur le sujet, il impose que les systèmes d'IA et leurs différentes applications soient examinés en fonction des risques qu'ils présentent pour les utilisateurs, et qui donnent lieu à différents niveaux d'obligation pour les fournisseurs. Alors que les applications et les systèmes associés à un risque « inacceptable[99] », comme les programmes de notation sociale mis en œuvre par le gouvernement chinois, sont interdits ; les « applications à haut risque[100] » telles que les procédures automatiques de filtrage des CV des candidats à l'emploi sont soumises à des exigences légales particulières – notamment en matière de qualité, de transparence, de supervision humaine, de gouvernance des données et de sécurité – et doivent être évaluées non seulement avant leur mise sur le marché mais aussi au cours de leur cycle de vie.

Une section moins importante porte sur les systèmes d'IA « à risque limité », qui sont soumis à des obligations de transparence plus souples : les développeurs et les déployeurs doivent seulement s'assurer que les utilisateurs ont conscience d'interagir avec des IA (par exemple dans le cas des *chatbots* ou des *deepfakes*). Les IA à usage général, qui peuvent poser un risque systémique, constituent une catégorie à part, ajoutée en 2023 du fait du succès des systèmes d'IA polyvalents comme ChatGPT : ceux-ci sont assortis d'obligations supplémentaires et d'une évaluation de sécurité spécifique, notamment sur les risques systémiques associés.

En revanche, les usages qui ne sont pas explicitement prohibés ou répertoriés comme présentant un risque élevé échappent largement à la réglementation. L'*AI Act* n'aborde pas non plus l'utilisation de l'IA dans des applications militaires, car ce domaine politique est réservé aux gouvernements nationaux dans le système européen. Cette catégorisation des risques et des obligations spécifiques pour les applications qui en découlent doit néanmoins assurer que les systèmes d'IA utilisés au sein de l'UE sont sûrs, transparents, traçables, non discriminatoires et respectueux de l'environnement.

98. Proposés par des experts indépendants nommés par le Bureau de l'IA, les principaux aspects de ce code comprennent des détails sur la transparence et l'application des règles relatives au droit d'auteur pour les fournisseurs de modèles d'IA à usage général, ainsi qu'une taxonomie des risques systémiques, des méthodes d'évaluation des risques et des mesures d'atténuation pour les fournisseurs de modèles avancés d'IA à usage général susceptibles de présenter des risques systémiques.
99. Ce qualificatif concerne notamment les applications de l'IA dont il est estimé qu'elles portent atteinte aux valeurs et aux droits fondamentaux de l'Union européenne, tels que la dignité humaine, la démocratie ou l'État de droit.
100. Sont ainsi définis les systèmes d'IA utilisés dans les secteurs de la santé, de l'éducation, du recrutement, de la gestion d'infrastructures critiques, du maintien de l'ordre ou de la justice.

Les grands axes de l'*AI Act*



Cette approche « *risk-based* » pourrait servir de modèle pour développer des standards globaux fondés sur la transparence, la sécurité, la responsabilité et la protection des droits humains. L'objectif assumé de l'UE est ainsi de créer un nouvel « effet Bruxelles » en entreprenant la première régulation d'ampleur de l'IA tout en promouvant les normes et les valeurs européennes à l'échelle internationale[101]. C'est en effet ce qui s'est produit en 2019, quand a été promulgué le Règlement général de protection des données personnelles (RGPD) : peu après sa ratification, l'initiative *Data Free Flow with Trust*, dérivée du RGPD, était approuvée par les pays du G20 à Osaka (Japon), puis différents dispositifs directement inspirés de la réglementation européenne ont essaimé au Brésil, en Chine en Corée du Sud ou encore au Chili.

Parce que c'est l'initiative normative la plus aboutie à ce jour, l'*AI Act* pourrait de fait servir de point de départ aux discussions multilatérales et devenir un modèle pour les législations nationales. De plus, compte tenu de l'influence considérable dont elle jouit dans les négociations internationales du fait de ses parts de marché, l'UE est en position de force pour inciter d'autres pays à adopter des règles contraignantes en matière d'IA. Ainsi que le résume Jean-Rémi de Maistre, « l'*AI Act* peut se révéler être un véritable atout stratégique pour l'Europe[102] ». Comme le RGPD, il s'applique de manière extraterritoriale aux produits et services utilisés mis en circulation sur le marché européen mais ayant un impact sur les fournisseurs étrangers : il s'étend ainsi à toute entité fournissant dans l'Union un produit ou un service fondé sur l'IA.

Sa mise en œuvre dans l'ensemble de l'UE est cependant confrontée à des défis allant de l'harmonisation entre les États membres de l'UE à l'implication des parties prenantes telles que les entités gouvernementales et les fournisseurs, importateurs, utilisateurs et distributeurs de systèmes d'IA. Le rayonnement de ce règlement dépendra également de la manière dont les autres régions réagiront, en particulier les grandes puissances de l'IA comme les États-Unis et la Chine[103]. En l'absence d'alignement au niveau mondial, il pourrait en résulter un amalgame de réglementations, qui obscurcissent les efforts de mise en conformité.

Les négociations de l'*AI Act* n'ont pas non plus été bien vécues par certaines start-ups éminentes du domaine, au premier rang desquelles la « licorne » française Mistal AI, forte de ses quelque 6 milliards d'euros de valorisation atteints en quelques mois. Comme le dénonçait son fondateur Arthur Mensch sur X en novembre 2023, le règlement européen favoriserait

101. A. Bradford, *The Brussels Effect: How the European Union Rules the World*, Oxford, Oxford University Press, 2020.
102. J.-R. de Maistre, « Intelligence artificielle : "L'innovation ne doit pas se faire au sacrifice de l'éthique et de la sécurité" », *Le Monde*, 11 septembre 2024.
103. R. Csernatoni, « The AI Governance Arms Race: From Summit Pageantry to Progress? », Carnegie Endowment for International Peace, 7 octobre 2024.

« les entreprises en place qui peuvent se permettre de faire face à de lourdes exigences de conformité », à savoir les géants du numérique et leur « armée d'avocats »[104].

La Chine en quête de leadership sur la gouvernance mondiale de l'IA

Tandis que l'*AI Act* était négocié dans les institutions européennes, la République populaire de Chine (RPC) a elle aussi amorcé une réflexion afin d'étoffer son approche de la régulation du secteur. Cette démarche s'inscrit dans une tendance plus large de volonté de refonte de la gouvernance internationale par la Chine, qui estime que ses intérêts ne sont pas correctement pris en compte par le système actuel[105]. L'objectif est aussi, dans le même temps, de maintenir un contrôle étroit du numérique à l'échelle nationale.

En 2021 puis en 2022, la Chine est ainsi devenue le premier pays à adopter des réglementations détaillées et strictes encadrant certaines des applications les plus répandues de l'IA, qui ont posé les bases d'un nouveau cadre de gouvernance. Cette structure politique évolutive est vouée à emporter des conséquences géopolitiques majeures, puisqu'elle influence à la fois la recherche exploratoire en IA, le fonctionnement de la deuxième économie mondiale, ainsi que des domaines tels que les grands modèles de langage en Afrique et les véhicules autonomes en Europe[106]. Le 2 juillet 2024, Pékin a encore annoncé vouloir édicter plus de cinquante nouvelles normes au sujet de l'IA jusqu'en 2026[107].

Depuis 2017, la Chine entend s'imposer comme le leader global de l'IA à l'horizon 2030. Cet objectif stratégique s'appuie sur un ambitieux plan de développement pour les nouvelles générations d'IA, en particulier en ce qui concerne l'IA générative. Ce projet s'accompagne de financements massifs : en mai 2024, l'agence Reuters rapportait ainsi que Big Fund, un fonds d'investissement spécialisé dans les semi-conducteurs et soutenu par l'État chinois, entamait sa troisième phase d'investissement, avec une enveloppe de 344 milliards de yuans, soit environ 44,12 milliards d'euros[108].

Dans le domaine normatif également, les décideurs politiques chinois ont insisté sur leur volonté d'être les premiers à agir, afin de s'offrir un leadership mondial en matière de gouvernance de l'IA. Depuis 2021,

104. C. Auffray, « Pour Arthur Mensch (Mistral AI), l'*AI Act* se trompe de cible », *ZDNet*, 22 mai 2024.
105. S. Kastner, M. Pearson et C. Rector, *China's Strategic Multilateralism: Investing in Global Governance*, Cambridge, Cambridge University Press, 2019.
106. M. Sheehan, « Tracing the Roots of China's AI Regulations », Carnegie Endowment for International Peace, février 2024.
107. J.-R. de Maistre, « Intelligence artificielle : "L'innovation ne doit pas se faire au sacrifice de l'éthique et de la sécurité" », *op. cit.*
108. « China Sets Up Third Fund with $47.5 Bln to Boost Semiconductor Sector », Reuters, 27 mai 2024.

plusieurs textes de lois sont ainsi venus encadrer l'innovation chinoise. Toutefois, à la différence de l'UE, qui a adopté une approche holistique et fondée sur l'atténuation des risques et le respect des valeurs, la Chine se penche sur des applications spécifiques de l'IA de manière séquentielle, à travers un ensemble de règles imposant de nouvelles obligations – d'abord dans le domaine des recommandations algorithmiques puis au sujet des techniques utilisées pour générer les contenus synthétiques (ou *deepfakes*)[109].

Ces règles imposent notamment aux fournisseurs d'indiquer quels contenus sont générés par l'IA et de veiller à ce que ces contenus ne violent pas les « droits à l'image » des personnes ou ne nuisent pas à « l'image de la nation ». Ensemble, ces deux réglementations ont également donné lieu à la création d'un registre des algorithmes, qui est devenu la pierre angulaire du régime chinois de gouvernance de l'IA.

En juillet 2023, la Chine a par ailleurs franchi une étape majeure dans la régulation de l'IA générative en adoptant un cadre réglementaire dédié à ces services[110]. Celui-ci a pour objectif de concilier progrès technologique et exigences de sécurité, tout en favorisant la croissance du secteur industriel. Ces mesures s'appliquent aux services d'IA générative proposés au public en Chine, que le fournisseur soit basé en Chine ou à l'étranger. De ce fait, les mesures relatives à l'IA générative ont un effet extraterritorial au même titre que l'*AI Act*.

Comme l'avance le chercheur Matt Sheehan, contrairement à une idée reçue, le régime de gouvernance de l'IA en Chine n'est pas le fruit d'un processus vertical. S'il arrive que le président Xi Jinping ou d'autres hauts responsables du Parti communiste chinois (PCC) donnent des orientations sur les priorités politiques, ils ne sont pas les principaux acteurs de l'élaboration des réglementations chinoises en matière d'IA. Celles-ci résultent plutôt d'une démarche dynamique et itérative, menée par un ensemble d'acteurs issus tant de l'intérieur que de l'extérieur du gouvernement chinois. Ces acteurs comprennent des bureaucrates de niveau intermédiaire, des universitaires, des représentants de la *tech*, des journalistes et des chercheurs des entreprises du numérique. Grâce à un mélange de plaidoyer public, de débat intellectuel, d'ateliers techniques et de querelles bureaucratiques, ces acteurs ont posé les fondations des réglementations actuelles et futures de la Chine en matière d'IA[111]. Il s'avère également ces efforts normatifs ont été principalement motivés par les craintes de déstabilisation politique liée à l'IA, qui poussent le PCC à encadrer fermement le secteur privé.

109. M. Sheehan, « Tracing the Roots of China's AI Regulations », *op. cit.*
110. M. Rochefort, « La Chine s'apprête à réglementer l'IA générative », *Siècle Digital*, 11 juillet 2023.
111. M. Sheehan, « Tracing the Roots of China's AI Regulations », *op. cit.*

Il semble par ailleurs que la Chine partage les préoccupations occidentales en matière de risques posés par l'IA[112]. Un nombre croissant de documents de recherche, de déclarations publiques et de documents gouvernementaux suggèrent que la sécurité de l'IA devient un sujet de plus en plus important en Chine, ce qui justifie à la fois des investissements techniques massifs et des interventions réglementaires. L'écho rencontré par ce thème s'est d'abord développé au sein de la technocratie chinoise, avant de gagner les cercles politiques les plus éminents du pays. En juillet 2024, alors que Shanghai accueillait une conférence mondiale sur l'IA, le PCC a notamment publié en juillet 2024 un document d'orientation qui appelle à la création de « systèmes de surveillance pour garantir la sécurité de l'intelligence artificielle ».

Sur le plan international, la Chine se démarque par ses efforts pour peser sur la normalisation de l'IA, à plusieurs niveaux. Entre 2014 et 2023, elle a ainsi déposé plus de 60 % des brevets en IA générative dans le monde, soit trois fois plus que les États-Unis, même s'ils ne parviennent pas véritablement à s'imposer comme des références mondiales[113]. Pékin s'illustre également par sa participation active au sein des instances internationales de normalisation, en particulier à l'Union internationale des télécommunications. Depuis octobre 2023, la Chine a également lancé sa propre *Global AI Governance Intitiative* et œuvre activement au développement capacitaire en matière d'IA, notamment dans le cadre des BRCIS+.

Compte tenu du désinvestissement des États-Unis des affaires multilatérales consécutif à l'investiture de Donald Trump, la Chine pourrait être tentée de reprendre le récit occidental à ses propres fins. Elle poursuivrait alors ses efforts de « *capacity building* » auprès des pays du Sud, notamment en ce qui concerne les transferts de technologies et l'accès aux marchés émergents. Dans le même temps, elle renforcerait la promotion de ses propres standards technologiques à des fins d'influence, tout en cherchant à capter toujours plus de marchés.

Une régulation menacée aux États-Unis

Parallèlement à ces processus en Europe et en Chine, les États-Unis ont imaginé leurs propres feuilles de route pour déterminer et tempérer les menaces associées aux technologies d'IA. Bien que le récit américain du « jeu à somme nulle » dans sa compétition avec la Chine limite la possibilité d'une régulation ambitieuse aux États-Unis, un phénomène d'inflation législative tend à se mettre en place, du fait de rapports de force internes. Ces initiatives visent en effet à apporter une réponse aux électeurs tout en

112. M. Sheehan, « China's Views on AI Safety Are Changing – Quickly », Carnegie Endowment for International Peace, août 2024.
113. « Generative Artificial Intelligence », *Patent Landscape Report*, World Intellectual Property Organization, juillet 2024.

réaffirmant le rôle de régulateur des institutions en place[114]. Le sénateur américain Chuck Schumer déclarait ainsi en avril 2023 qu'il attendait des États-Unis qu'ils ne laissent pas la Chine « prendre la première position en termes d'innovation, ni écrire le Code de la route[115] » en matière d'IA.

L'approche de l'administration Biden en matière de gouvernance de l'IA s'est d'abord traduite par le *Blueprint for an AI Bill of Rights*, publié en octobre 2022. Ce cadre mettait en avant cinq principes clés pour guider le développement responsable de ces technologies, notamment au regard de la protection contre les biais algorithmiques et la prise en compte de la vie privée. Il n'a toutefois pas fait l'objet d'adoption à ce jour et sert uniquement de guide non contraignant aux agences et entreprises impliquées dans le développement ou le déploiement de systèmes d'IA.

Puis, en collaboration avec les secteurs privé et public, le National Institute of Standards and Technology (NIST) a mis au point le AI Risk Management Framework afin de mieux gérer les risques associés à l'IA – que ce soit pour les individus, les organisations ou la société. Conçu pour une utilisation volontaire, il vise à améliorer la capacité à intégrer des considérations de fiabilité dans la conception, le développement, l'utilisation et l'évaluation des produits, services et systèmes d'IA aux États-Unis.

En septembre 2023, l'administration Biden-Harris a également obtenu des engagements volontaires de la part d'entreprises leaders du secteur pour un développement sécurisé et transparent des technologies d'IA[116]. Ces engagements consistent notamment à s'assurer que les produits sont sûrs avant de les présenter au public, à mettre en place des systèmes axés sur la sécurité et la confiance du public.

Un mois plus tard, Washington présentait son premier décret de grande envergure sur l'IA, l'*Executive Order* 14110 du président Biden, instaurant une panoplie de normes, de mesures de sécurité, de protection de la vie privée et de contrôle pour le développement et l'utilisation d'une « intelligence artificielle sûre, sécurisée et digne de confiance ». Ce décret entendait notamment élucider des problématiques liées à l'équité et aux droits civiques, en abordant également des applications spécifiques de l'IA. Il se traduisait par des évaluations robustes, fiables, reproductibles et normalisées des systèmes d'IA, ainsi que par la mise en œuvre de politiques publiques, d'institutions et de mécanismes permettant de tester, de comprendre et d'atténuer les risques liés à ces systèmes avant qu'ils ne soient

114. R. Heath, « Exclusive: States Are Introducing 50 AI-related Bills per Week », *Axios*, 14 février 2024, disponible sur : www.axios.com.
115. D. Shepardson, « US Senate Leader Schumer Calls for AI Rules as ChatGPT Surges in Popularity », Reuters, 13 avril 2023.
116. « Fact-Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage Risks Posed by AI », Washington D.C., Maison-Blanche, 21 juillet 2023, disponible sur : www.whitehouse.gov.

utilisés. Comme pour la Chine et l'UE, l'*Executive Order* 14110 contenait des clauses ayant une portée extraterritoriale[117].

Ce décret présidentiel a marqué une étape cruciale dans la définition de la réglementation et de la responsabilité en matière d'IA dans de multiples secteurs, en mettant l'accent sur une approche « pan-gouvernementale » pour traiter à la fois les opportunités et les risques associés à l'IA. Il constitue l'un des efforts les plus ambitieux du gouvernement américain pour encadrer le développement et l'utilisation de cette technologie, et visait à positionner les États-Unis en tant que leader dans l'adoption de pratiques sûres, éthiques et responsables en matière d'IA. L'*Executive Order* chargeait les agences fédérales de s'atteler à plusieurs enjeux majeurs : la gestion des modèles d'IA à double usage, l'instauration de protocoles de test rigoureux pour les systèmes à haut risque, la mise en place de mécanismes de responsabilisation, la défense des droits civils et la garantie de transparence à chaque étape du cycle de vie de l'IA. Ces mesures devaient permettre de réduire les risques pour la sécurité, de protéger les valeurs démocratiques et de renforcer la confiance du public dans ce secteur en constante évolution.

L'avenir de cette initiative demeure toutefois incertain à l'aube du mandat Trump II. Alors que le président avait annoncé en juillet 2024 sa volonté de révoquer le décret s'il était élu[118], il a tenu parole quelques heures à peine après son investiture le 20 janvier 2025[119], avant d'annoncer le lendemain le lancement du projet *Stargate* à hauteur de 500 milliards de dollars sur quatre ans pour « bâtir les infrastructures physiques et virtuelles pour porter la prochaine génération d'IA[120] ».

Une grande partie des efforts de réglementation de l'IA reposant sur le travail des agences fédérales, les principes établis pourraient perdurer même après l'abrogation du décret, permettant aux États-Unis de conserver leur influence dans la définition de normes en la matière. Néanmoins, le rôle de plus en plus prépondérant occupé par les magnats de la *tech* américains dans la nouvelle administration – à l'image de l'investisseur David Sacks, propulsé « White House AI & Crypto Czar[121] » – met en péril les velléités d'encadrement de l'IA, d'autant qu'ils défendent un agenda résolument anti-

117. En particulier une clause qui impose au département du Commerce de « proposer des réglementations obligeant les fournisseurs américains de services d'accès à l'Internet à soumettre un rapport au Comité de surveillance lorsqu'une personne étrangère effectue des transactions pour former un grand modèle d'IA doté de capacités potentielles qui pourraient être utilisées dans des activités cybernétiques malveillantes. »

118. Au-delà de l'échelle fédérale émergent des initiatives locales, à l'image de deux projets de loi introduits en Californie pour encadrer le recours à l'IA dans le domaine de l'emploi.

119. D. Shepardson, « Trump Revokes Biden Executive Order on Addressing AI Risks », Reuters, 21 janvier 2025.

120. S. Holland, « Trump Announces Private-sector $500 Billion Investment in AI Infrastructure », Reuters, 22 janvier 2025.

121. A. Chow, « What Trump's New AI and Crypto Czar David Sacks Means For the Tech Industry », *Time*, 10 décembre 2024.

régulation. Quoi qu'il en soit, l'impact à long terme des efforts de l'administration Biden–Harris dépendra de la capacité des décideurs politiques à suivre le rythme effréné des avancées en IA, tout en préservant un équilibre entre soutien à l'innovation et confiance publique[122].

Des initiatives multilatérales en « patchwork »

À la course à l'innovation technologique s'ajoute une vive compétition entre les États, dont beaucoup rêvent de devenir la première puissance normative de l'IA. Les trois grands blocs s'efforcent donc de prendre le leadership en matière de gouvernance, afin de faire advenir des réglementations compatibles avec leurs ambitions nationales, et susceptibles de freiner au passage leurs compétiteurs. Cette régulation en « patchwork » fait cependant craindre une balkanisation de la gouvernance, qui verrait l'adoption fragmentée de normes concurrentes voire contradictoires, et qui irait de fait à l'encontre de l'ambition de régulation universelle d'une technologie qui l'est tout autant.

C'est la raison pour laquelle de nombreuses alternatives multilatérales ont vu le jour au cours des dernières années. D'abord, le Partenariat mondial pour l'IA (PMIA), créé en 2019 sous l'impulsion du Canada et de la France, réunit 29 membres (28 États et l'UE) dans le but de renforcer la coopération multi-acteurs Nord-Sud pour faire émerger des IA reflétant les valeurs démocratiques et répondant aux défis mondiaux. Le PMIA prône également l'utilisation responsable et centrée sur l'humain de l'IA, respectant les droits humains et les libertés fondamentales. Hébergé à l'Organisation de coopération et de développement économique (OCDE), il s'appuie également sur deux centres de recherche en IA, l'Institut national de recherche en sciences et technologies du numérique (INRIA) en France et le Centre d'expertise international de Montréal en intelligence artificielle (CEIMIA) au Canada, afin de favoriser la collaboration et le partage de connaissances entre la société civile, les gouvernements et le monde universitaire. Les travaux du PMIA s'articulent autour de quatre groupes d'experts : *Responsible AI*, *Data Governance*, *Future of Work* et *Innovation and Commercialization*. Du fait des limites géographiques de ses membres et de l'absence d'accords contraignants, le PMIA fait toutefois l'objet de critiques régulières pour son manque d'universalisme, de représentativité – il ne compte que sept pays du Sud aux côtés de l'OCDE et de douze pays de l'UE – et d'efficacité.

122. A. Kein, C. Kerry, C. Radsch, M. MacCarthy, S. Friedley et N. Turner Lee, « 1 Year Later, How Has the White House AI Executive Order Delivered on Its Promises? », Brookings Institution, 4 novembre 2024.

Pays membres du PMIA

En 2019, l'OCDE a pour sa part mis au point des principes relatifs à l'IA[123]. Adoptés par 47 pays, ils promeuvent une IA innovante et digne de confiance, respectueuse des droits humains et des valeurs démocratiques. Leur actualisation en 2024, qui vise à prendre en compte les nouveaux développements technologiques et politiques, témoigne d'ailleurs de la nécessité d'une régulation « *future proof* » afin que les normes établies demeurent robustes et adaptées à leur objectif en dépit de l'évolution des technologies et de l'émergence de nouvelles applications.

Pays adhérents aux principes de l'OCDE sur l'IA

123. Ils se composent plus précisément de cinq principes fondés sur des valeurs et de cinq recommandations qui doivent servir d'orientations pratiques et flexibles aux décideurs politiques et aux acteurs de l'IA.

Ces principes établis par l'OCDE bénéficient d'un rayonnement certain : les États membres se fondent sur ces lignes directrices pour établir leurs édifices normatifs en matière d'IA, jetant ainsi les bases d'une interopérabilité mondiale entre les différentes juridictions. À titre d'exemple, l'UE, le Conseil de l'Europe, les États-Unis et même les Nations unies utilisent la définition d'un système d'IA posée par l'OCDE dans leurs cadres législatifs et réglementaires et dans leurs orientations politiques.

Tableau comparatif des principes de l'UNESCO et de l'OCDE sur l'Intelligence artificielle

Principes et recommandations	UNESCO	OCDE
Equité, valeurs humaines et non-discrimination	✓	✓
Transparence et explicabilité	✓	✓
Sécurité, sûreté, robustesse	✓	✓
Responsabilité	✓	✓
Coopération internationale et gouvernance de l'IA	✓	✓
Proportionnalité et innocuité	✓	✗
Droit au respect de la vie privée et protection des données	✓	✗
Surveillance et décisions humaines	✓	✗
Durabilité	✓	✗
Sensibilisation et éducation	✓	✗
Croissance inclusive, développement durable et bien-être	✗	✓
Investir dans la R&D en matière d'IA	✗	✓
Façonner un cadre d'action favorable à l'IA	✗	✓
Favoriser l'instauration d'un écosystème numérique pour l'IA	✗	✓
Renforcer les capacités humaines et préparer la transformation du marché du travail	✗	✓

Texte en gras : recommandation de l'OCDE à destination des décideurs politiques

En novembre 2021, l'UNESCO à quant à elle proposé son tout premier standard mondial sur l'IA, la « Recommandation sur l'éthique de l'Intelligence artificielle », adoptée par l'ensemble de ses 193 États membres. Au cœur de cette recommandation se retrouve la protection de la dignité et des droits humains, par le renforcement de principes fondamentaux tels que la transparence, l'équité, la responsabilité et le contrôle des systèmes d'IA. De plus, l'UNESCO y a identifié de nombreux domaines d'action stratégique, supposés permettre aux décideurs politiques de traduire les valeurs en actes, notamment sur les questions de gouvernance des données, d'environnement, de genre, d'éducation, de recherche, de santé et de bien-être social. Bien que non contraignant,

cette recommandation a rencontré un certain succès en février 2024, lorsque huit grandes entreprises de la *tech*, dont le géant Microsoft, se sont officiellement engagées à en respecter les valeurs et les principes à chaque étape de la conception et du déploiement de leurs systèmes d'IA[124].

Lors du G7 de mai 2023, organisé sous présidence japonaise, a également été lancé le Processus d'Hiroshima », dans le but de définir les grands principes d'une gouvernance de l'IA générative et plus largement des modèles d'IA avancés (« *frontier AI* »). Pilotée par les Japonais, cette initiative se compose de plusieurs groupes de travail créés spécifiquement pour aborder ces questions. La France, le Canada, les États-Unis, le Royaume-Uni, l'Italie, l'Allemagne et la Commission européenne participent aux travaux, tandis que l'OCDE et le PMIA interviennent en tant qu'organisations invitées.

En octobre 2023, les pays du G7 se sont également accordés sur un code de conduite volontaire à destination des entreprises développant des systèmes d'IA avancés. Les onze points de ce code visent à « promouvoir mondialement une IA sûre et digne de confiance » et à « contribuer à saisir les avantages et à répondre aux risques et défis apportés par ces technologies ». Les entreprises y sont incitées à prendre des mesures adaptées pour identifier, évaluer et réduire les risques tout au long du cycle de vie d'un système d'IA, et à remédier aux éventuels incidents sur des produits d'IA déjà commercialisés. Le code les enjoints également à mettre en place des contrôles de sécurité robustes et à faire preuve de transparence sur leurs capacités ou les obstacles qu'elles rencontrent. Le G20 n'est pas en reste dans ces efforts de régulation de l'IA tous azimuts. Dans le cadre de la présidence brésilienne de 2024, a ainsi été mise en exergue une volonté d'aborder l'IA sous l'angle du développement économique, de l'accès à l'IA et à ses infrastructures pour les pays en développement.

Une autre tendance observée dans les efforts multilatéraux d'encadrement des technologies d'IA tient à l'émergence d'une forme de « diplomatie des sommets », qui voit différents États chercher à s'affirmer comme chef de file de la gouvernance mondiale en organisant de vastes rencontres internationales. Sous l'impulsion du Premier ministre Rishi Sunak, qui souhaitait en faire son legs politique, le Royaume-Uni a ainsi organisé en novembre 2023 le tout premier sommet mondial sur l'IA, l'Artificial Intelligence Safety Summit, à Bletchley Park. Cet événement, qui a réuni des acteurs multilatéraux et les diverses parties prenantes sur les risques liés à l'IA générative, a non seulement marqué un succès diplomatique pour Londres, mais également initié une dynamique de

124. P. Rioux, « Intelligence artificielle : 8 géants de la tech s'engagent à appliquer la recommandation éthique de l'Unesco », *La Dépêche*, 7 février 2024.

coopération encourageante, même si la place occupée dans les échanges par les grands acteurs privés a été critiquée[125].

Avant même ce sommet, Rishi Sunak avait d'ailleurs annoncé la création d'un AI Safety Institute, conçu pour être reproduit dans les autres États participants au sommet – comme l'ont notamment fait les États-Unis au sein du NIST. Les 20 et 21 novembre 2024, les premiers AI Safety Institutes et les bureaux mandatés par le gouvernement d'Australie, du Canada, de la Commission européenne, de France, du Japon, du Kenya, de la République de Corée, de Singapour, du Royaume-Uni et des États-Unis se sont d'ailleurs retrouvés à San Francisco pour la première réunion du Réseau international des instituts de sécurité de l'IA[126]. Ces instituts pourraient alors dépasser leur mission initiale et devenir une source de normes internationales, dans une démarche plus *bottom-up* et axée sur les enjeux techniques concrets[127].

Au terme de ce premier AI Safety Summit, 28 États dont la Chine, les États-Unis et l'UE se sont par ailleurs accordés sur une déclaration commune, dite « Déclaration de Bletchley », qui témoigne d'une volonté de coopération pour établir un cadre normatif garantissant que l'IA est développée et utilisée de manière responsable et fiable dans le monde entier. Dans la lignée du Royaume-Uni, la Corée du Sud et la France ont à leur tour organisé des sommets mondiaux sur l'IA. Le sommet de Séoul de mai 2024, à nouveau axé sur les enjeux de sécurité, a notamment donné lieu à un engagement de seize entreprises majeures du secteur (en Chine, en Corée du Sud, aux Émirats arabes unis et aux États-Unis) en faveur de l'atténuation des risques[128].

Pour s'imposer comme l'une des capitales mondiales de l'IA et se distinguer des sommets précédents, Paris adopte une approche résolument positive et optimiste. La France a ainsi décidé de mettre l'accent sur l'action, qui donne son nom au sommet des 10 et 11 février 2025, pour mettre en valeur l'écosystème national de l'IA en même temps que sa propre approche des enjeux. « Nous allons mettre en évidence les risques de l'IA, déjà bien évoqués à Londres, en novembre 2023 et à Séoul en mai, mais aussi les opportunités et les bénéfices de cette technologie », explique l'Élysée[129].

125. M.-F. Cuéllar, « The UK AI Safety Summit Opened a New Chapter in AI Diplomacy », Carnegie Endowment for International Peace, 9 novembre 2023.
126. Dans le cadre du Trade and Technology Council mis en place par Joe Biden et Ursula von der Leyen entre 2021 et 2024, les AI Safety Institutes ont d'ailleurs été mobilisés pour étendre la coopération pour une IA sûre, fiable et responsable.
127. D. Milmo, « UK's AI Safety Institute "Needs to Set Standards Rather Than Do Testing" », *The Guardian*, 11 février 2024.
128. Parmi lesquelles Amazon, Google, Microsoft, Meta, Mistral AI ou encore OpenAI. Lire « Seoul Declaration for Safe Innovative and Inclusive AI: AI Seoul Summit 2024 », Department for Science, Innovation and Technology, disponible sur : www.gov.uk.
129. A. Piquard, « Le sommet de Paris vise à agir contre les risques et surtout pour les bénéfices de l'IA », *Le Monde*, 9 décembre 2024.

Le pari français est aussi d'inclure davantage la société civile, d'élargir le spectre de problématiques abordées[130] et de réconcilier les enjeux de sécurité et le soutien à l'entrepreneuriat. Cet événement pourrait également ouvrir la voie à des innovations institutionnelles, comme la création d'une « Organisation mondiale de l'IA[131] » ou d'une « Organisation mondiale des données[132] », afin d'imaginer des solutions collectives à ces défis globaux et à instaurer un cadre de coopération constructif autour de ces enjeux.

Les pays du Sud semblent suivre l'exemple de leurs homologues occidentaux : le Rwanda a ainsi annoncé vouloir organiser prochainement le premier sommet international sur l'IA en Afrique[133]. L'Inde est également particulièrement investie tant dans le développement que dans la gouvernance de l'IA[134], comme en témoignent sa présidence du PMIA en 2024 et sa co-présidence de l'AI Action Summit de février 2025. Toutefois, cette politique de sommets internationaux peut avoir un effet pernicieux tant, comme le déplorait le chercheur Stuart Russell en juillet 2024, « l'IA est vue comme un véhicule du nationalisme économique[135] », et est devenue un enjeu de *nation branding* à part entière.

Le cadre onusien semble à cet égard le plus pertinent pour éviter les logiques de « *forum shopping* » et concevoir une régulation à la fois robuste et universelle de ces technologies. Les efforts en la matière se dessinent petit à petit : le 18 juillet 2023, le conseil de sécurité des Nations unies a, pour la première fois, tenu une réunion dédiée à l'IA. « La nature même de la technologie – transfrontalière dans sa structure et son application – nécessite une approche mondiale », concluait également le rapport final du groupe d'experts nommé par Antonio Guterres sur le sujet[136]. De plus, en mars 2024, l'Assemblée générale des Nations unies a adopté – par 193 votes « pour » – une résolution visant à établir des règles internationales encadrant les usages de l'IA, pour « combler le fossé numérique » et limiter les risques[137].

130. En l'espèce, le sommet s'articule autour de cinq volets : l'intérêt général, le travail, la culture, la confiance et la gouvernance mondiale de l'IA.
131. Comme l'a notamment suggéré le rapport de la Commission nationale de l'IA. Lire « IA : notre ambition pour la France », Commission de l'Intelligence artificielle, mars 2024.
132. I. Bremmer, « Why We Need a World Data Organization. Now », *GZERO*, 25 novembre 2019.
133. « Rwanda Announces Plans to Host Inaugural High-Level Summit on AI in Africa », Rwanda Centre for the Fourth Industrial Revolution, 18 janvier 2024.
134. En mars 2024, le pays s'est doté d'une stratégie sur l'IA (*IndiaAI Mission*), avec un budget d'1,2 milliard de dollars sur cinq ans dédié à sept piliers, dont la gouvernance fait partie. Le gouvernement veut en effet promouvoir une IA sûre et de confiance, en utilisant les applications développées par les pouvoirs publics comme des modèles de bonne conduite et en développant des outils techniques permettant de garantir un certain cadre éthique aux applications de l'IA.
135. A. Piquard, « À l'approche du sommet de Paris, les militants inquiets quant à la "sécurité de l'IA" cherchent à se faire entendre », *Le Monde*, 11 septembre 2024.
136. Le Comité consultatif de haut niveau des Nations unies sur l'IA, créé le 26 octobre 2023, est composé de 39 experts du monde entier. Lire « La nécessité d'une règlementation de l'IA est « irréfutable » assurent des experts de l'ONU », ONU Info, 20 septembre 2024, disponible sur : news.un.org.
137. Cette résolution exclut le domaine de la défense. De manière générale, la question de la gouvernance de l'IA militaire se concentre essentiellement sur l'objet des armes autonomes et fait l'objet d'une

Les actions concrètes associées à ce rapport et à cette résolution demeurent cependant floues, d'autant que l'avis du groupe d'experts est consultatif et que la résolution n'est pas engageante sur le plan juridique.

Les progrès modestes du PMIA – ou, plus récemment, des sommets sur la sécurité de l'IA – démontrent également la difficulté de converger vers une régulation mondiale dans un monde profondément fragmenté. Dans les faits, il ne suffit pas que des États adhèrent à une nouvelle institution internationale pour que les normes se mettent en place d'autant que, dans le domaine de l'IA, les acteurs non étatiques s'avèrent incontournables.

Les alternatives émanant d'acteurs non étatiques

L'entreprise de gouvernance internationale de l'IA ne saurait en effet être complète sans la participation active de ceux qui en conçoivent les technologies, à savoir les grandes plateformes américaines du numérique. Comme le soulignait le Secrétaire général des Nations unies, Antonio Guterres, à Davos en janvier 2024 : « Il est urgent que les gouvernements travaillent avec les entreprises technologiques à l'élaboration de cadres de gestion des risques, de suivi et d'atténuation des préjudices futurs, au regard du développement actuel de l'IA. » Lesdites entreprises sont toutefois loin d'être parfaitement coopératives, parce qu'elles tendent à voir la régulation comme un obstacle à l'innovation. Pour autant, leurs efforts de *lobbying* dans les enceintes de négociation internationales sont tels qu'elles ont désormais la possibilité d'orienter les grandes lignes de la gouvernance nationale et internationale de l'IA[138].

Les géants du numérique américains adoptent ainsi une approche agressive pour influencer les choix normatifs à la Maison-Blanche, au sein de l'administration, du Congrès, et même en Europe. Leur objectif est d'orienter la régulation vers les risques à long terme, au détriment des enjeux immédiats, afin de se ménager une plus grande liberté d'action[139]. Cette démarche met en lumière une contradiction flagrante chez des figures comme Elon Musk ou Sam Altman qui, bien qu'ils portent des discours alarmistes sur l'Intelligence artificielle générale (AGI) et prétendent appeler à une régulation stricte, participent activement à cette course technologique tout en cherchant à saper les cadres réglementaires, ce qui leur permet de reporter leur propre responsabilité sur les régulateurs. Même lorsqu'ils font

conversation internationale séparée, dans le cadre de la Convention sur certaines armes classiques (CCAC) à Genève. Toutefois, alors que le groupe d'experts gouvernementaux doté d'un mandat de discussion sur le cadre normatif et opérationnel à apporter à ces technologies travaille depuis plus de dix ans, les efforts de régulation sont dans l'impasse. Pour une ample analyse de cette entreprise normative, lire L. de Roucy-Rochegonde, *La Guerre à l'ère de l'Intelligence artificielle*, Paris, PUF, 2024.

138. J. Tallberg *et al.*, « AI Regulation in the European Union: Examining Non-State Actors Preferences », *Business and Politics*, vol. 26, 2024.

139. B. Pajot, « Les risques de l'IA : enjeux discursifs d'une technologie stratégique », *op. cit.*

mine de craindre une véritable catastrophe, comme ils l'ont prétendu dans une lettre ouverte de mai 2023 appelant à un moratoire sur l'IA générative, ce n'est que pour mieux freiner leurs compétiteurs et rattraper leur retard technologique.

Ainsi que s'en inquiétait la chercheuse Courtney Radsch en novembre 2024, il y a fort à parier que ni des approches aussi ambitieuses que l'*AI Act* européen ni même des initiatives plus nuancées à l'image de *l'Executive Order* de Joe Bien ne parviennent à « atténuer la puissance monopolistique d'une poignée de géants de la technologie[140] ». Alors que plusieurs patrons de la Silicon Valley ont prêté allégeance à Donald Trump avant même son investiture le 20 janvier 2025[141], un agenda fortement anti-régulation et même anti-Europe semble poindre. Le 7 janvier 2025, Mark Zuckerberg, le CEO de Meta, Instagram et WhatsApp, a ainsi annoncé mettre fin à ses partenariats de *fact-checking* avec plusieurs grands médias américains et internationaux, au profit d'une politique de modération des contenus fondée sur le principe des « *community notes* » qui existent déjà sur X. Dans son communiqué, il s'en est également pris violemment à l'Europe où, dit-il, « un nombre croissant de lois [...] rendent difficile la construction de projets innovants ». Il a du même coup annoncé sa volonté de travailler avec Donald Trump à « repousser les gouvernements du monde entier qui s'en prennent aux entreprises américaines[142] ».

Ce soutien au président nouvellement investi semble amorcer une offensive contre l'Europe, ciblant les amendes et taxes imposées par la Commission européenne – perçues comme un moyen de pénaliser les entreprises américaines – ainsi que des réglementations jugées trop restrictives pour permettre l'innovation. Bien que les géants de la *tech* se soient conformés aux règles sur la protection des données, ils paraissent désormais déterminés à opposer un refus catégorique à la régulation en matière d'IA qu'ils considèrent comme excessive et qui est vue comme un enjeu crucial pour les États-Unis.

Cette antienne sur l'antagonisme supposé entre innovation et régulation n'est pas nouvelle dans la bouche des patrons des *Big Tech*[143]. En septembre 2024, déjà, une trentaine d'entreprises du numérique publiaient une lettre ouverte dénonçant une Europe « devenue moins compétitive et innovante que d'autres régions et risquant aujourd'hui de reculer encore dans l'ère de l'Intelligence artificielle, en raison de décisions

140. A. Kein, C. Kerry, C. Radsch, M. MacCarthy, S. Friedley et N. Turner Lee, « 1 Year Later, How Has the White House AI Executive Order Delivered on Its Promises? », *op. cit.*

141. A. Leparmentier, « Après Musk et Bezos...Zuckerberg : la *tech* en ordre de marche derrière Trump », *Le Monde*, 8 janvier 2025.

142. D. Leloup et A. Piquard, « La fin des partenariats de *fact-checking* chez Meta, un revirement symbolique », *Le Monde*, 7 janvier 2025.

143. Sur les accusations américaines contre l'Europe et le caractère crucial du débat innovation versus régulation, lire M. Velliet, « Souveraineté numérique : politiques européennes, dilemmes américains », *Notes de l'Ifri*, Ifri, mars 2023, disponible sur : www.ifri.org.

de régulation incohérentes ». Mark Zuckerberg allait même plus loin, en suspendant le lancement dans l'UE de son assistant d'IA, Meta AI, sur Instagram et Facebook[144]. Dans son communiqué, le groupe enjoignait Bruxelles à arrêter de « rejeter le progrès [...] et de regarder le reste du monde construire des technologies auxquelles les Européens n'auront pas accès ». Ce débat dépasse en réalité les enjeux sociaux, économiques et juridiques tant il s'inscrit dans une logique stratégique. Le narratif de la compétition technologique est exploité pour promouvoir une approche minimaliste de la régulation, présentée comme essentielle pour préserver l'innovation et conserver une longueur d'avance sur les rivaux internationaux.

Il est pourtant essentiel de souligner que l'hypothèse selon laquelle la régulation freinerait l'innovation et affaiblirait les États occidentaux sur le plan géopolitique n'a pas été démontrée. Cette idée soutenue par les grands acteurs américains du secteur prend souvent les atours d'une mise en garde contre le risque d'une supériorité technologique chinoise, qui n'est pas non plus prouvée. L'approche européenne tente justement de résoudre cette tension supposée entre innovation et régulation, et de montrer que l'une et l'autre sont complémentaires. De fait, en offrant un cadre juridique clair, en unifiant la réglementation des différents marchés nationaux à l'échelle de l'Europe et en permettant à d'autres acteurs que les seuls géants de la *tech* d'y accéder, les normes européennes s'avèrent plutôt propices à la concurrence et à l'innovation, en même temps qu'elles permettent de développer sereinement ces technologies.

Pour tenter de rassurer leurs utilisateurs sans pour autant avoir les mains liées par des règlements juridiquement contraignants, beaucoup d'entreprises se sont autosaisies des questions d'encadrement de l'IA en les ancrant dans le domaine de l'éthique. En septembre 2016, plusieurs grandes entreprises du numérique américaines avaient ainsi lancé un partenariat sur l'éthique de l'IA : le *Partnership on Artificial Intelligence to Benefit People and Society*[145]. Une telle prudence s'explique par leur perception d'un risque à l'image, alimenté par les actions d'ONG poursuivant des stratégies de « *naming and shaming* » destinées à entacher la réputation d'acteurs engagés dans des activités présentées comme nuisibles.

Cette tendance est parfois appelée « *ethics washing* ». Comme le *greenwashing*, cette pratique consiste à feindre une considération éthique pour améliorer la façon dont une personne ou une organisation est perçue. Les exemples de cette pratique sont particulièrement visibles dans de nombreux projets axés sur le développement et l'utilisation des nouvelles

144. A. Piquard et V. Malingre, « IA : Meta et Apple mettent la pression sur l'Union européenne, accusée de "rejeter le progrès" », *Le Monde*, 25 septembre 2024.
145. Dont faisaient notamment partie Google, Facebook, IBM, Microsoft et Amazon. Lire M. Tual, « Intelligence artificielle : les géants du Web lancent un partenariat sur l'éthique », *Le Monde*, 28 septembre 2016.

technologies. Dans leur article « Why Are We Failing at the Ethics of AI? », Anja Kaspersen et Wendell Wallach affirment ainsi :

> « Ces dernières années ont vu une prolifération d'initiatives sur l'éthique de l'intelligence artificielle (IA). Qu'elles soient formelles ou informelles, menées par des entreprises, des gouvernements, des organisations internationales et à but non lucratif, ces initiatives ont développé pléthore de principes et d'orientations pour soutenir l'utilisation responsable des systèmes d'IA et des technologies algorithmiques. Malgré ces efforts, peu d'entre elles sont parvenues à moduler réellement les effets de l'IA. »[146]

Les deux chercheurs décrivent également l'*ethics washing* dans le domaine de l'IA comme « créant un sentiment rassurant mais illusoire que les questions éthiques sont traitées de manière adéquate, afin de justifier la poursuite de systèmes », alors même que ceux-ci vont dans un sens opposé aux précautions en vigueur. L'approche par les questionnements éthiques présente aussi l'avantage de se traduire par des engagements non contraignants et éminemment subjectifs, contrairement à la réglementation qui s'accompagne nécessairement de mesures de surveillance et de vérification. De plus, du point de vue de la légitimité démocratique, le type d'entité qui adopte les réglementations en matière d'IA et les motifs pour lesquels ces entités décisionnelles sont dotées d'un tel pouvoir ont une grande importance[147].

Malgré ces déclarations d'intention, les véritables priorités des *Big Tech* peuvent être questionnées, puisqu'alors que les applications de l'IA se développent, les équipes consacrées aux questions d'éthique ou aux pratiques responsables de l'IA chez Meta, Google, Microsoft et Amazon voient leurs effectifs diminuer[148]. OpenAI a ainsi annoncé en mai 2024 la dissolution de son équipe chargée de la sécurité d'une potentielle superintelligence artificielle, réorientée vers le développement technologique[149].

146. A. Kaspersen et W. Wallach, « Why Are We Failing at the Ethics of AI? », Artificial Intelligence and Equality Inititative, Carnegie Council for Ethics in International Affairs, 10 novembre 2021.
147. E. Erman et M. Furendal, « Artificial Intelligence and the Political Legitimacy of Global Governance », *Political Studies*, vol. 72, n° 2, 2022.
148. G. de Winck et W. Oremus, « As AI Booms, Tech Firms Are Laying Off Their Ethicists », *The Washington Post*, 30 mars 2023, disponible sur : www.washingtonpost.com.
149. « OpenAI : dissolution de l'équipe chargée de la sécurité d'une potentielle superintelligence artificielle », *Le Monde*, 18 mai 2024.

Pour une gouvernance de l'IA inclusive et pérenne

Reste donc à savoir si ces initiatives pléthoriques donneront lieu à des engagements concrets, permettront de diluer les risques les plus préoccupants dans les années à venir et, surtout, si elles trouveront un écho au-delà de la sphère occidentale, alors que seuls 7 pays participent à toutes et que 119 ne font partie d'aucune. Bien que les acteurs publics comme privés aient commencé à travailler à la réduction des risques liés à l'IA, leurs approches demeurent dispersées. À mesure que les tentatives d'encadrement se multiplient dans le monde entier, il est essentiel de réfléchir à leur mise en cohérence, pour prévenir une fragmentation des cadres normatifs, qui pourrait entraîner une tension entre des modèles opposés voire incompatibles. Une coopération internationale plus forte est alors indispensable pour remédier à cet éparpillement, mettre en place des garde-fous robustes et bâtir une gouvernance de l'IA inclusive et pérenne.

L'inclusivité au service du consensus

Il est désormais clair qu'une gouvernance mondiale est cruciale pour coordonner les normes, empêcher l'exploitation des « zones grises » juridiques et prévenir le danger d'une course mondiale à l'IA dépourvue d'éthique. Les forums multilatéraux semblent le lieu privilégié pour harmoniser les diverses approches de la régulation. Pour autant, la construction d'un consensus international sur ces questions est loin d'être acquise. Cette entreprise s'avère au contraire éminemment délicate, à la fois parce qu'elle nécessite de concevoir un langage diplomatique sur des enjeux techniques inédits, parce que les États craignent un déclassement technologique et stratégique, et parce que ces débats interviennent dans une période de crise profonde du multilatéralisme[150], où sont contestés les grands accords de coopération internationaux.

Derrière les différences de priorités et de perspectives sur l'encadrement de l'IA se cachent en effet des intérêts nationaux bien compris. D'une part, les États tentent de défendre leurs entreprises, qui

150. Le multilatéralisme fait l'objet d'un débat à la fois politique et académique, parce qu'il est en même temps un instrument fonctionnel et une invention politique. Subissant de multiples crises (de vitalité, de fonctionnalité et d'universalité notamment), il est critiqué en tant que mode d'organisation (du fait de la politisation ou de la dépolitisation des relations internationales qu'il occasionne) mais aussi en tant que concept. Il est ici entendu au sens le plus littéral, en tant que forme majeure de l'action internationale. Pour une analyse substantielle de ces débats, lire J. Fernandez et J.-V. Holeindre, *Nations désunies ? La crise du multilatéralisme dans les relations internationales*, Paris, CNRS Éditions, 2022.

estiment que les efforts de régulation vont à l'encontre de leur capacité à rivaliser avec leurs concurrents. Cette tendance a été particulièrement visible dans les réticences françaises au cours des tractations sur l'*AI Act*. Au prétexte de la défense de l'innovation, Paris voulait en réalité assurer les possibilités de croissance de la prometteuse (et surtout française) Mistral AI, qui développe des modèles de langage *open source* et propriétaires pour des applications en IA générative.

Au-delà de la recherche d'un avantage concurrentiel, les parties prenantes veulent également imposer un modèle de valeurs sous-tendant le développement et le déploiement de l'IA, à l'image de l'UE qui insiste en particulier sur le respect des droits humains. Les pays du Sud, en revanche, se concentrent davantage sur les conséquences sociales et les inégalités économiques créées par ces technologies émergentes. Ils sont par conséquent partagés entre une vision de la régulation comme nécessaire pour atténuer les externalités négatives dont ils sont souvent les premières victimes ; ou comme frein à leur émancipation par la technologie. La gouvernance internationale de l'IA doit alors s'accompagner d'efforts systématiques pour améliorer l'accès à ces innovations : les économies en développement doivent bénéficier de leur potentiel, afin de combler le fossé numérique au lieu de l'aggraver. Le sommet de Paris sur l'IA entend aller dans ce sens, puisque devraient y être annoncées les adhésions de nouveaux pays du Sud au PMIA, ainsi que la création d'une fondation à Paris pour abaisser les barrières à l'entrée et renforcer l'accès aux communs numériques. Une levée de fonds de 2,5 milliards d'euros est en cours dans ce but[151].

Plus largement, la recherche d'un consensus ne pourra aboutir que si la parole est donnée à tous les acteurs. Comme le notait le rapport du groupe d'experts de l'ONU sur l'IA : « L'équité exige qu'un plus grand nombre de voix jouent un rôle significatif dans les décisions touchant à la bonne gouvernance des technologies qui nous affectent[152]. » Les représentants de la société civile ont eux aussi un rôle à jouer dans ces réflexions. D'ailleurs, ils sont peut-être les mieux placés pour contrebalancer le poids exorbitant des géants du numérique dans ce paysage normatif. Dans le domaine, certes un peu différent, de la maîtrise des armements, plusieurs grandes conventions internationales ont vu le jour à l'issue d'efforts de la société civile pour endiguer les conséquences humanitaires catastrophiques de certaines armes, et ce en dépit des réticences étatiques ou du *lobbying* intense des grands industriels de la défense[153].

151. Ce qui est aussi un moyen de ne pas laisser à la Chine le monopole du *capacity building*. Entretien de recherche de l'auteure avec un conseiller à la présidence de la République.
152. « La nécessité d'une règlementation de l'IA est "irréfutable" assurent des experts de l'ONU », *op. cit.*
153. C'est ce que l'on appelle le désarmement humanitaire, caractérisé par l'irruption de la société civile dans les enceintes de négociations diplomatiques, et qui ont déjà permis l'interdiction des armes à lasers aveuglantes, des mines antipersonnel et des armes à sous-munitions.

Enfin, il est indispensable d'adopter une approche plus intégrée et cohérente des différentes initiatives composant le « complexe de régimes[154] » de l'IA. Le Sommet de l'Avenir 2024, qui s'est tenu les 22 et 23 septembre 2024 à New York, a ainsi été l'occasion pour l'ONU et l'OCDE d'annoncer une nouvelle collaboration pour renforcer la gouvernance mondiale de l'IA. Dans son annonce, Amandeep Singh Gill, l'Envoyé du Secrétaire général des Nations unies pour les technologies, a souligné l'importance cruciale de la collaboration entre ces deux entités : « le développement rapide des technologies d'IA et l'ampleur de leur impact nécessitent une collaboration renforcée, et en temps réel, entre les divers écosystèmes de politiques publiques[155] ».

Cette nouvelle initiative doit mettre à profit la complémentarité de l'ONU – qui jouit d'un rayonnement universel – et de l'OCDE – dont les compétences techniques et analytiques sont reconnues – pour aider les gouvernements à agir rapidement et efficacement en réponse à l'ensemble des enjeux liés à l'IA. L'objectif affiché est de mettre en place des mécanismes de gouvernance mondiale solides, en collaboration avec des parties prenantes majeures, notamment des universités de premier plan et des chercheurs de renommée internationale.

Dans la même veine, l'AI Action Summit de Paris a pour objectif de faire émerger un consensus international sur un socle commun pour la gouvernance de l'IA. Pour atteindre cet objectif, un processus inclusif de co-construction est prévu avec plus de 70 parties prenantes, incluant des États, des organisations internationales, des chercheurs, des entreprises et des ONG. En réunissant des représentants de ces différents domaines et en instaurant des mécanismes de consultation étendus, l'ambition est de faire émerger une vision partagée et des positions communes, reflétant au mieux les préférences et les intérêts collectifs.

Quel organe de régulation ?

Il ne s'agit toutefois pas seulement de parvenir à une adoption commune de normes, mais aussi de garantir leur application effective à l'échelle internationale. Dans ce domaine, les institutions multilatérales jouent un rôle clé, en établissant des canaux de communication, en conciliant les approches divergentes entre pays et en renforçant la transparence et la coopération. À l'image de l'Organisation mondiale du commerce, de la Cour internationale de Justice ou de la Cour pénale internationale, elles ont pour mission de faire respecter les règles adoptées et de résoudre les différends qui pourraient émerger dans un domaine où les intérêts nationaux sont

154. R. Csernatoni, « Charting the Geopolitics and European Governance of Artificial Intelligence », Carnegie Europe, 6 mars 2024.
155. « L'OCDE et les Nations Unies annoncent les prochaines étapes de leur collaboration sur l'intelligence artificielle », Communiqué de presse, OCDE, 22 septembre 2024, disponible sur : www.oecd.org.

souvent en compétition. Certains plaident donc pour la création d'un nouvel organe *ad hoc*. Quelles innovations institutionnelles seraient alors nécessaires pour mieux coordonner la régulation de l'IA ?

Plusieurs précédents sont présentés comme pouvant servir de modèle. Premièrement l'IA peut, au même titre que le nucléaire, être définie comme une technologie « polyvalente » ou « d'application générale », du fait de trois grandes caractéristiques : son immense potentiel d'amélioration continue, son omniprésence dans de très nombreux secteurs industriels, et sa complémentarité avec d'autres technologies. Par conséquent, certaines leçons peuvent être retenues des organes d'encadrement du nucléaire.

En ce qui concerne la recherche d'abord, plusieurs associations ainsi que l'École polytechnique fédérale de Zurich militent pour la création d'un « CERN de l'IA ». Située non loin de Genève, l'organisation européenne pour la recherche nucléaire est en effet le plus grand centre de physique des particules au monde. C'est une infrastructure de recherche publique, en source ouverte et internationale. De la même manière que le CERN est doté d'accélérateurs de particules, le centre de recherche proposé par le Large-Scale Artificial Intelligence Open Network (Laion) disposerait de machines équipées de quelque 100 000 accélérateurs[156] (par exemple des processeurs graphiques – GPU). Celles-ci seraient pilotées par les États participant et pourraient être utilisées par des chercheurs internationaux – dépendamment des niveaux d'autorisation exigés, comme dans les laboratoires de recherche biologiques. Tous les résultats seraient ensuite rendus publics.

Une telle infrastructure présenterait l'avantage d'émanciper la recherche en IA de l'emprise des grandes multinationales. Elle ne répondrait toutefois pas au besoin de mécanismes de surveillance et de vérification indispensables au bon fonctionnement d'un organe de régulation. Toujours selon l'exemple du nucléaire, certains proposent alors un modèle similaire à celui de l'Agence internationale de l'énergie atomique (AIEA).

L'AIEA, basée à Vienne, a été créée en 1957 sous l'égide de l'ONU et promeut l'utilisation sûre, sécurisée et pacifique des technologies nucléaires, tout en surveillant d'éventuelles violations du Traité sur la non-prolifération des armes nucléaires de 1968 (TNP). Forte de ses 178 États membres, de sa présence dans le monde entier et de l'adhésion des grandes puissances nucléaires, elle œuvre au développement de l'énergie nucléaire pour la production d'électricité, et à la limitation de ses applications militaires.

156. R. Karayan, « IA générative : une association allemande milite pour une recherche ouverte sur le modèle du CERN », *L'Usine digitale*, 6 avril 2023.

Parce que l'IA est une technologie duale, une telle perspective peut sembler pertinente. En effet, il existe de très nombreux usages pacifiques et même bénéfiques des techniques d'IA, qui ne doivent pas être suspendus sous prétexte d'endiguer les menaces, tout aussi prégnantes, qui y sont associées[157]. L'intérêt d'une agence capable de se saisir de ces deux volets est donc considérable, d'autant plus que les discussions sur la gouvernance internationale de l'IA « civile » et « militaire » sont pour l'instant hermétiquement séparées, ce qui a peu de sens compte tenu des possibles détournements d'applications « civiles » à des fins malveillantes.

Cette proposition s'avère d'ailleurs très populaire, à la fois auprès de décideurs politiques tels que Rishi Sunak, de patrons de la *tech* comme Sam Altman et d'Antonio Guterres. Pour autant, elle est loin d'être aisée à mettre en œuvre. En effet, si l'AIEA a vu le jour en 1957, il a fallu attendre l'entrée en vigueur du TNP en 1970 pour que l'agence soit en mesure d'effectivement contrôler les programmes d'armement nucléaire des pays participants et de faire respecter les normes de sécurité. Or, les discussions actuelles sur la gouvernance de l'IA ne tiennent pas suffisamment compte du rôle essentiel des traités. Tout à leur impatience de créer de nouvelles institutions internationales, les dirigeants politiques tendent à oublier que les capacités de contrôle ne peuvent émerger que d'engagements contraignants de la part des États et n'être garanties que si elles sont assorties de mécanismes de sanction en cas d'infraction.

Dans un ouvrage sur l'AIEA, l'historienne Elisabeth Roehrlich met ainsi en lumière deux éléments essentiels qui ont rendu le travail de l'AIEA sur les garanties nucléaires efficace : les accords juridiques liant l'agence et ses États membres, et les outils techniques permettant de contrôler le respect de ces accords[158]. Une gouvernance robuste de l'IA doit de la même manière se traduire à la fois par des normes nouvelles et par des ressources et des capacités techniques permettant d'assurer leur mise en œuvre.

Les organismes internationaux de régulation ne peuvent mener leur mission à bien que lorsque leur mandat est concret et que des règles claires sont établies et applicables aux entreprises et aux gouvernements. Il revient alors aux décideurs politiques de commencer par définir les conditions préalables et le contenu de ces lois avant de mettre en place des agences chargées de leur application. L'IA, en raison de son développement rapide, de son opacité et de ses évolutions constantes, se distingue fondamentalement des technologies précédentes et exige des formes inédites de contrôle

157. Il en allait d'ailleurs de même pour ce qui concerne les armes chimiques. C'est la raison pour laquelle l'Organisation pour l'interdiction des armes chimiques a non seulement pour mission de détruire les arsenaux existants et les installations militaires qui y sont liées dans le monde entier, mais aussi de surveiller certaines activités de l'industrie chimique pour éviter le risque de diversion militaire, et promouvoir la coopération entre États pour une utilisation pacifique de la chimie.
158. E. Roehrlich, *Inspectors for Peace: A History of the International Atomic Energy Agency*, Baltimore, Johns Hopkins University Press, 2022.

international. Plutôt que de se laisser intimider par l'ampleur de ce défi, les régulateurs devraient y voir une opportunité d'innovation et de créativité dans la conception de cadres réglementaires adaptés.

Dans un autre registre, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) est souvent pris en exemple de ce qui pourrait être imaginé pour l'IA. Rishi Sunak, ainsi que d'autres personnalités comme l'ancien PDG de Google, Eric Schmidt, se disent inspirés par le modèle du GIEC, qui compile les recherches scientifiques sur le changement climatique et organise les sommets annuels de la Conférence des parties (COP). Avant même que le Royaume-Uni n'organise le premier sommet sur la sécurité de l'IA, les plans pour ce « GIEC de l'IA » mettaient cependant en avant un mandat clair : l'organisme ne serait pas chargé de formuler des recommandations politiques. Son rôle consisterait plutôt à synthétiser régulièrement les recherches en IA, à identifier les préoccupations communes et à présenter des options politiques, sans pour autant donner de directives explicites. Cette approche limitée ne semble donc pas non plus ouvrir la voie à un traité contraignant, qui offrirait des garanties solides et limiterait réellement l'influence des entreprises technologiques. Un « GIEC de l'IA », en tant que groupe de recherche indépendant, ne serait pas inutile, en particulier compte tenu de l'opacité des informations fournies par les entreprises sur la technologie et de l'importance de s'appuyer sur un consensus scientifique pour établir des règles. Cependant, faciliter la recherche n'est qu'une étape intermédiaire vers l'élaboration de normes, sans lesquelles une gouvernance efficace de l'IA ne saurait voir le jour.

S'inspirer de modèles comme l'AIEA ou le GIEC pour la gouvernance de l'IA risque enfin de négliger la spécificité et la nouveauté des défis que pose cette technologie. Contrairement au nucléaire, dont la maîtrise relève principalement des gouvernements, les capacités en matière d'IA sont concentrées entre les mains de quelques entreprises qui commercialisent activement leurs produits. Elles sont aussi beaucoup moins coûteuses et difficiles à fabriquer que ne l'est, par exemple, le fait d'enrichir de l'uranium en vue de concevoir une arme nucléaire. De ce fait, le potentiel de diffusion des technologies d'IA, y compris à des acteurs non étatiques et malveillants, est immense.

La traduction des grands principes en termes techniques

Il ne suffit pas de créer un organe de régulation pour garantir le respect des principes qui émergent au sujet de l'IA. Se pose aussi la question de l'opérationnalisation des grands principes qui voient le jour. Bien qu'ils soient en mesure de guider les acteurs de l'IA dans leurs efforts pour développer une IA digne de confiance et de fournir aux décideurs politiques des recommandations pour des politiques efficaces, ils demeurent généraux et laissent une large marge d'interprétation. Pour que les garde-fous soient

robustes, il est nécessaire d'harmoniser les standards d'une part et de transposer techniquement les principes sur lesquels les acteurs se sont accordés. Les organismes internationaux de standardisation sont à ce titre cruciaux pour faire advenir une gouvernance de l'IA qui dépasse les vœux pieux et qui puisse véritablement contraindre les géants du numérique américains ou chinois.

Les normes techniques sont en effet essentielles pour définir les paramètres des systèmes d'IA, qu'il s'agisse des architectures de référence de base, des exigences en matière de sécurité et d'éthique ou du fonctionnement technique d'applications spécifiques dans un large éventail de domaines, notamment les soins de santé, l'éducation, la fabrication de pointe, l'énergie et l'agriculture. Or, dans leurs efforts pour maîtriser et canaliser le développement de l'IA, tant la Chine que l'UE se sont tournées vers l'établissement de normes techniques pour atténuer les risques et atteindre des objectifs politiques généraux[159].

Cette transposition donne toutefois lieu à une vive compétition, dans ce que Paul Timmers a appelé la « géopolitique de la normalisation[160] », c'est-à-dire l'élaboration des normes techniques prévalant à l'usage des systèmes informatiques. Celle-ci se joue au niveau national, régional – en particulier européen – et international. Or, la gouvernance des organisations de normalisation est atypique. Ainsi, pour siéger au European Telecommunications Standards Institute (ETSI), « il suffit de payer et plus le montant est élevé, plus le nombre de votes augmente[161] ». Par conséquent, là où les grandes entreprises du numérique chinoises et américaines sont sur-représentées, les entreprises européennes sont peu nombreuses, car le prix du ticket d'entrée est très élevé.

Le Comité européen de normalisation et le comité européen de normalisation électronique (CEN-CENELEC) adoptent quant à eux les normes telles qu'édictées par l'Organisation internationale de normalisation (ISO) et la Commission électrotechnique internationale (IEC), dans lesquelles siège un représentant par État, mais dont les places sont, là aussi, prédatées par les géants du numérique. Ainsi, les représentants les plus actifs sur la normalisation de l'IA sont actuellement des salariés de Microsoft, d'IBM, de Google et de Huawei, puisque la représentation se fait par nationalité, indépendamment de l'entreprise d'appartenance. En 2022, le chef de délégation irlandais et la cheffe de délégation autrichienne du comité de l'ISO sur la normalisation de l'IA étaient salariés de Huawei, tandis que les chefs de délégation britannique et allemand étaient salariés de Microsoft. Par ailleurs, dans une fiche de poste proposée par Huawei à l'un de ces experts,

159. C. Wang et Y. Yin, « China Launches Global AI Governance Initiative, Offering an Open Approach in Contrast to US Blockade », *Global Times*, 18 octobre 2023.
160. P. Timmers, « Géopolitique de la normalisation », *Le Grand Continent*, 2 juin 2020, disponible sur : https://legrandcontinent.eu.
161. Entretien de recherche de l'autrice avec un expert de la normalisation.

était inscrit parmi les objectifs celui de (*sic*) : « contrer la législation européenne[162] ».

De ce fait, les normes européennes et mondiales sur les systèmes d'IA sont élaborées en premier lieu par des entreprises privées, avec des biais certains. Au-delà du problème évident de souveraineté posé par ce fonctionnement, il est possible de voir des stratégies étatiques poindre derrière ces acteurs privés. Comme l'affirme Paul Timmers au sujet de la Chine :

> « Il s'agit d'une stratégie délibérée du gouvernement chinois pour fixer les règles du jeu dans les nouveaux domaines des technologies de l'information et de la communication (TIC) et pour rompre avec les normes passées, qui ont été largement déterminées par les États-Unis et l'Europe. »[163]

Dans le cadre des travaux de mise en application de l'*AI Act*, la Commission européenne a, par exemple, fourni un certain nombre de directives, dont les décrets d'application sont *in fine* définis par les travaux de normalisation. Lorsque la Commission énonce par exemple que l'IA doit être « robuste », c'est la traduction technique de cet adjectif dans les enceintes de standardisation internationales qui crée la règle. Puis la « robustesse » devient un standard s'appliquant aux systèmes d'IA. Ceux qui rédigent les normes techniques disposent alors d'un pouvoir extraordinaire, ce que certains pays comme les États-Unis et la Chine ont bien compris. Comme le soulignait Thierry Breton en février 2022 : « Les normes techniques revêtent une importance stratégique. La souveraineté technologique de l'Europe, sa capacité à réduire ses dépendances et la protection des valeurs de l'UE dépendront de notre capacité à définir des normes à l'échelle mondiale ».

De plus, au-delà de la compétition sur la normalisation technique pour projeter de la puissance et des valeurs, il n'est pas clair où ces principes doivent prendre corps : dans les procédures ou dans les objectifs[164]. L'approche dominante consiste ainsi à se concentrer sur les résultats – la production normative – en partant d'un problème potentiel créé par l'IA pour identifier des principes de gouvernance susceptibles de minimiser les risques ou de rendre plus probable un résultat souhaité[165]. À l'inverse, il est possible de mettre l'accent sur les procédures, en accordant plus d'importance à la manière dont les processus de gouvernance se déroulent.

Ainsi, le principe de justice peut être compris à la fois comme une valeur procédurale (en veillant à l'équité et à la représentativité des instances de gouvernance) ou comme un résultat distributif (en s'assurant

162. *Ibid.*
163. P. Timmers, « Géopolitique de la normalisation », *op. cit.*
164. J. Tallberg *et al.*, « The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research », *op. cit.*
165. A. Dafoe, *AI Governance: A Research Agenda*, *op. cit.*

que, techniquement, les systèmes d'IA n'alimentent pas des injustices). Par exemple, il est probable que l'automatisation induite par l'IA ne fasse perdre leurs emplois à certaines populations, qui supporteraient alors une part disproportionnée des externalités négatives de la technologie sans que celle-ci ne soit compensée par l'accès à ses avantages[166]. Se concentrer uniquement sur la justice en tant que valeur procédurale reviendrait à négliger ces effets distributifs créés par la diffusion des systèmes d'IA.

Alors que sont édictés les grands principes devant prévaloir au développement et au déploiement de l'IA – confiance, robustesse, explicabilité, responsabilité, sûreté, sécurité, inclusivité, durabilité, respect des valeurs démocratiques et des droits humains, etc.[167] – il est indispensable de réfléchir à leurs déclinaisons techniques, procédurale et normative.

La nécessaire articulation avec les régulations nationales

Parce que les acteurs privés au cœur de l'écosystème de l'IA peuvent opérer dans plusieurs juridictions nationales, les efforts de régulation doivent également être transfrontaliers. Ce n'est qu'en introduisant des règles communes que les États pourront s'assurer que ces entreprises sont exposées à des environnements réglementaires similaires. Seule une telle approche peut favoriser le développement de l'IA dans le monde et réduire les incitations pour les entreprises à se tourner vers des pays où la régulation est plus laxiste.

Pour autant, la gouvernance ne pourra être efficace à l'échelle internationale que si elle est articulée avec les régulations nationales. La promotion d'institutions pour établir des normes et des standards et les faire respecter sans établir en même temps des règles nationales est au mieux naïve, et au pire délibérément intéressée – elle peut traduire une forme de « fausse *compliance* » visant à freiner des compétiteurs. La transposition des normes internationales dans les approches nationales est en effet essentielle pour garantir l'interopérabilité des standards entre les différentes juridictions et éviter d'ajouter à la cacophonie normative. C'est la raison pour laquelle l'AI Action Summit entend s'appuyer sur « la convergence des standards et des politiques publiques propres à l'IA » pour bâtir un cadre de gouvernance robuste.

166. E. Erman et M. Furendal, « Artificial Intelligence and the Political Legitimacy of Global Governance », *op. cit.*

167. Les principes de l'OCDE sur l'IA sont ainsi la croissance inclusive, le développement durable et le bien-être ; le respect de l'état de droit, des droits humains et des valeurs démocratiques (en particulier équité et vie privée) ; la transparence et l'explicabilité ; la robustesse, la sûreté et la sécurité ; et enfin la responsabilité.

Toutefois, la vigilance est une fois de plus de mise quant à l'entrisme des géants du numérique dans les efforts nationaux de réglementation. Eric Schmidt a par exemple investi des sommes considérables dans des start-ups et des entreprises de recherche en IA, tout en conseillant le gouvernement américain sur la politique à adopter dans le domaine – en mettant sans surprise l'accent sur l'autonomie des entreprises. Ce risque manifeste de conflits d'intérêts rend nécessaire la mise en place des garde-fous juridiquement applicables et donnant la priorité à l'intérêt public, plutôt qu'à des normes peu contraignantes au service des résultats financiers des *Big Tech*[168].

Une dernière difficulté tient au fait que les applications de l'IA sont susceptibles d'enfreindre des lois préexistantes, sur des sujets extrêmement variés. De ce fait, les organismes de surveillance en construction doivent être en mesure de faire respecter les lois *antitrust*, la non-discrimination ou encore les lois sur la propriété intellectuelle déjà en vigueur, ce qui rend là aussi la coordination avec les réglementations à l'échelle nationale indispensable.

Vers une gouvernance « *future proof* »

Au-delà de l'enjeu de l'adoption et de l'application de nouvelles normes encadrant l'IA se pose aussi la question de leur potentielle obsolescence, compte tenu des avancées technologiques parfois fulgurantes dans ce domaine. Traditionnellement sont distingués quatre modes d'élaboration des normes[169].

D'abord, les règles existantes peuvent être réinterprétées pour couvrir également l'IA[170]. Par exemple, de nombreuses voix s'élèvent pour exiger que les principes de distinction, de proportionnalité et de précaution du droit international humanitaire soient étendus, par le biais d'une réinterprétation, pour s'appliquer aux systèmes d'armes létales autonomes, sans pour autant changer le texte juridique original.

La nouvelle réglementation de l'IA peut aussi être le fruit d'ajouts aux règles existantes. Ainsi, dans le domaine des véhicules autonomes, des dispositions relatives à l'IA ont été ajoutées à la Convention de Vienne sur la circulation routière de 1968, par le biais d'un amendement en 2015[171].

168. M. Schaake, « The Premature Quest for International AI Cooperation », *op. cit.*
169. J. Tallberg *et al.*, « The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research », *op. cit.*
170. M. Maas, *Artificial Intelligence Governance under Change:Foundations, Facets, Frameworks*, Copenhague, University of Copenhagen, 2021.
171. M. Kunz et S. Ó hÉigeartaigh, « Artificial Intelligence and Robotization » *in* R. Geiss et N. Melzer (dir.), *The Oxford Handbook on the International Law of Global Security*, Oxford, Oxford University Press, 2020.

La régulation peut enfin prendre la forme d'un cadre entièrement nouveau, soit du fait d'une nouvelle pratique étatique qui se transforme en droit international coutumier, soit par le biais d'une innovation normative tel qu'un nouvel acte juridique ou d'un nouveau traité[172] – comme l'est par exemple l'*AI Act*.

Ces différents modes de création de la norme ne prémunissent cependant pas contre le risque de voir des innovations s'inscrire dans des vides juridiques, compte tenu de la vitesse, de l'ampleur, et des incertitudes associées au développement de l'IA. De plus, les États utilisent souvent de nouvelles technologies bien avant de se mettre d'accord sur des règles spécifiques pour réguler leur usage[173]. Entre-temps, ils développent des manières de les employer et bâtissent des conceptions de ce qu'est un usage « approprié » de ces technologies, qui transforment les normes existantes.

La gouvernance de l'IA risque alors d'être constamment en retard sur les avancées technologiques. Personne ne peut prédire les capacités futures de l'IA : c'est la raison pour laquelle les politiques et les institutions qui la régissent doivent être conçues de manière flexible et adaptative, pour résister à l'épreuve du temps et de l'innovation. L'article 97 de l'*AI Act* permet ainsi à la Commission d'adopter des actes délégués pour mettre à jour la réglementation afin de tenir compte de l'évolution technologique. Dans la même veine, la Convention-cadre du Conseil de l'Europe sur l'IA a pour but d'éviter les vides juridiques qui pourraient résulter d'avancées technologiques trop rapides. Ainsi, afin de résister au temps, la Convention-cadre ne régule pas la technologie et est essentiellement neutre sur le plan technologique.

172. M. Maas, *Artificial Intelligence Governance under Change: Foundations, Facets, Frameworks*, *op. cit.*
173. R. Alcala et E. Talbot Jansen, *The Impact of Emerging Technologies on the Law of Armed Conflict*, New York, Oxford University Press, 2019.

Conclusion

Alors que l'IA s'impose chaque jour davantage comme une technologie incontournable, de nombreuses questions fondamentales demeurent sans réponse. Comment protéger efficacement la vie privée des utilisateurs face à une collecte massive de données ? Quels mécanismes peuvent garantir la transparence et l'équité des algorithmes lorsqu'ils influencent des décisions cruciales, comme l'accès à l'emploi ou aux services publics ? Surtout, comment éviter les dérives qui permettent à l'IA de générer des contenus dangereux ou d'armer des systèmes autonomes ? Ces défis illustrent l'urgence de définir des lignes directrices communes, afin d'encadrer les usages de cette technologie de manière éthique et responsable.

Sans une gouvernance mondiale cohérente, la dynamique de compétition technologique pourrait reproduire les erreurs du passé. À l'instar de la course à l'armement nucléaire au XX^e^ siècle, les États et les entreprises privées risquent de privilégier la rivalité économique et géopolitique au détriment de la sécurité collective[174]. Pour endiguer cette trajectoire, les décideurs politiques doivent non seulement démêler les bénéfices potentiels des risques associés à l'IA, mais aussi encourager un développement qui maximise les premiers, tout en atténuant les seconds.

Le sommet de Paris de février 2025 et les initiatives portées par des institutions comme les Nations unies, le G7 ou l'OCDE marquent des étapes importantes vers la clarification de la gouvernance mondiale de l'IA. Cependant, une approche véritablement coordonnée, fondée sur les risques et assurant l'interopérabilité entre les différents cadres réglementaires, reste à bâtir. Cette démarche est essentielle pour prévenir les abus, réduire les inégalités entre les régions du monde et garantir que les bénéfices de l'IA soient partagés équitablement.

Au fur et à mesure que la communauté internationale progresse dans sa compréhension de ces nouvelles techonologies, l'accent doit être mis sur des actions concrètes plutôt que sur des gestes symboliques. Si les sommets, les codes de conduites, les règlements et les déclarations ont mis en lumière l'importance de la gouvernance de l'IA, des engagements plus contraignants sont nécessaires pour entreprendre un véritable changement. Pour aller au-delà des vœux pieux ou des déclarations d'intention, il est donc impératif de concevoir des mécanismes de vérification et de sanction en cas d'infraction. L'AI Action Summit de Paris pourrait de ce point de vue

174. A. Dafoe, *AI Governance: A Research Agenda*, *op. cit.*

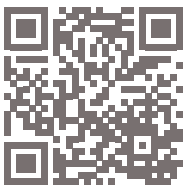
constituer un *momentum* et permettre la mise au point d'un « accord de Paris sur l'IA ».

Bien sûr, le multilatéralisme n'a pas le vent en poupe, encore plus alors que les États-Unis de Donal Trump II dénoncent leurs engagements internationaux les uns après les autres (Organisation mondiale de la santé, accord de Paris sur le climat…). Il faut néanmoins y voir une opportunité de revigorer un multilatéralisme moribond, en s'emparant du défi on ne peut plus universel qu'est l'encadrement de l'IA. La Convention-cadre du Conseil de l'Europe sur l'IA constitue à cet égard une avancée prometteuse. Contrairement à d'autres initiatives, ce texte est juridiquement contraignant et offre un modèle de structuration de la gouvernance de l'IA à l'échelle mondiale. C'est aussi une étape importante puisque c'est l'une des premières fois que les États-Unis et l'Ue s'alignent officiellement sur la réglementation de l'IA[175].

Au Sommet de l'avenir de septembre 2024, les dirigeants des 193 États des Nations unies ont par ailleurs adopté à l'unanimité le « Pacte de l'avenir », qui a pour objectif de réinventer le système multilatéral, ainsi que le « Pacte numérique mondial », qui doit leur permettre de se saisir des défis à long terme en la matière. Ont également été créés deux nouveaux mécanismes internationaux : un panel scientifique international indépendant sur l'IA et un dialogue mondial sur la gouvernance de l'IA. Celui-ci pourrait être la première pierre de l'édifice normatif complexe qui doit encadrer cette technologie.

Pour mettre en œuvre cet agenda ambitieux, un organe international devrait être établi, afin d'harmoniser les différentes initiatives et répartir entre elles les compétences (éthiques, sécuritaires, sociétales, scientifiques, techniques, commerciales, etc.), afin d'éviter les redondances et les contradictions. Il ne s'agit en aucun cas de repartir de zéro, mais bien de mettre en musique ce qui existe déjà. Aucune véritable gouvernance mondiale de l'IA ne saurait émerger sans l'existence d'un tel « chef d'orchestre ».

175. R. Csernatoni, « The AI Governance Arms Race: From Summit Pageantry to Progress? », *op. cit.*



27 rue de la Procession 75740 Paris cedex 15 – France

Ifri.org