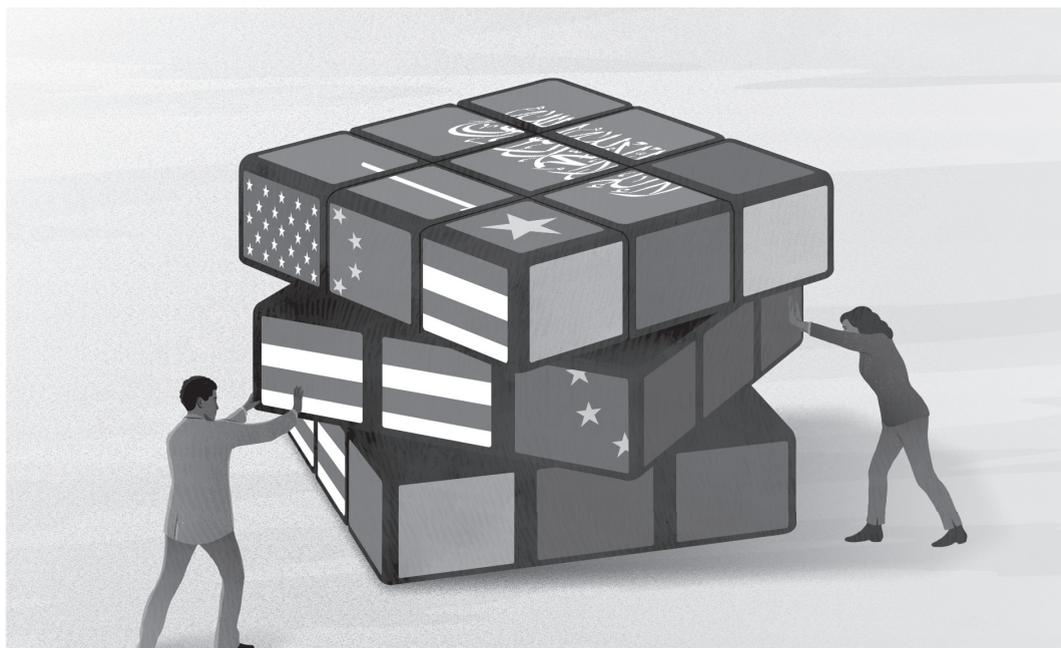


# ramses 2025

SOUS LA DIRECTION DE THIERRY DE MONTBRIAL ET DOMINIQUE DAVID

ifri



## ENTRE PUISSANCES ET IMPUISSANCE

- MOYEN-ORIENT : LA RECOMPOSITION SANS FIN
- ÉTATS-UNIS : L'EMPIRE INQUIÈTE...
- UNION EUROPÉENNE : UN DESTIN CONTRARIÉ ?

DUNOD

# L'Intelligence artificielle : quelle gouvernance ?

## Régulation éclatée ou unitaire ?

Les développements de l'Intelligence artificielle (IA) induisent à la fois espoirs et craintes. De multiples forums internationaux tentent d'organiser une coopération en vue d'une régulation qui n'aurait de sens que globale. Il faut désormais dépasser le stade des principes généraux pour qu'elle puisse se concrétiser.

**D**epuis 2023 et le succès commercial de l'agent conversationnel ChatGPT, les innovations de l'IA font l'objet d'une couverture médiatique considérable. Si les promesses de l'IA n'en finissent pas d'être mises en avant – meilleurs soins de santé, transports plus sûrs et plus propres, gains de productivité, énergie moins chère et plus durable... –, les risques qui les accompagnent sont tout aussi régulièrement énoncés.

Début mai 2023, l'un des pionniers de la recherche en IA, Geoffrey Hinton, présentait sa démission à Google, alertant sur les menaces potentielles de son champ de recherche. Dans un article publié dans le *New York Times*, il avançait que « les futures versions de cette technologie pourraient être un risque pour l'humanité ». Il s'y inquiétait de la sophistication croissante des techniques d'IA et du risque qu'elles ne tombent entre de mauvaises mains. Désormais, confiait-il, « une part de moi-même regrette l'œuvre de ma vie ».

Parce que cette technologie tend à se démocratiser et à se diffuser de manière exponentielle, sa réglementation ne peut être pensée qu'à l'échelle mondiale. Pourtant, aux obstacles classiques du multilatéralisme s'ajoutent les spécificités techniques de l'IA, qui rendent sa régulation d'autant plus complexe.

### Des risques inhérents à l'IA

Ces dernières années ont vu les alertes sur les menaces sous-jacentes de l'IA se multiplier. Les déclarations de Vladimir Poutine annonçant en 2017 que « celui qui deviendra leader dans l'IA sera le maître du monde », ou d'Elon Musk estimant en 2023 que « l'IA est bien plus dangereuse que l'arme nucléaire » et « causera probablement une troisième guerre mondiale », ont marqué l'imaginaire collectif.

Sans aller jusqu'à une vision aussi dystopique, les risques inhérents à l'IA sont nombreux, du contrôle social à la surveillance de masse et la désinformation, en

passant par le renforcement des inégalités et des discriminations. L'enjeu de la propriété intellectuelle est également revenu fréquemment sur le devant de la scène au cours des derniers mois, l'IA générative et les LLM (*Large Language Models*) type ChatGPT faisant voler en éclat l'acceptation traditionnelle du droit d'auteur.

Face à ces risques bien réels, les acteurs de l'IA et les régulateurs se sont mis en ordre de marche, mais avancent en ordre dispersé. En résulte une cacophonie réglementaire qui peine à s'harmoniser.

## Une balkanisation de la gouvernance

Les efforts de régulation de l'IA se sont accentués à partir d'avril 2021, lorsque la Commission européenne a proposé le plan initial de son *AI Act*. Finalement adopté au printemps 2024, c'est le premier cadre réglementaire international pour cette technologie. Il impose que les systèmes d'IA et leurs différentes applications soient analysés et classés en fonction des risques qu'ils présentent pour les utilisateurs, donnant ensuite lieu à différents niveaux de réglementation pour les fournisseurs. L'objectif est que les systèmes d'IA utilisés au sein de l'Union européenne (UE) soient sûrs, transparents, traçables, non discriminatoires et respectueux de l'environnement.

Alors que l'*AI Act* était négocié dans les institutions européennes, la Chine a elle aussi amorcé une réflexion sur son approche de la régulation du secteur. Les décideurs politiques chinois ont insisté sur leur volonté d'être les premiers à agir, afin d'offrir un leadership mondial en matière de gouvernance de cette technologie. Contrairement à 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 – des « recommandations algorithmiques » aux techniques utilisées pour générer les fameux *deepfakes* qui alertent tant les démocraties libérales.

Parallèlement à ces processus qui ont lieu en Europe et en Chine, le Royaume-Uni et les États-Unis ont imaginé leurs propres feuilles de route pour déterminer et tempérer les menaces associées aux technologies d'IA. Sous l'impulsion du Premier ministre britannique Rishi Sunak, qui souhaite en faire son legs politique, Londres organisait en novembre 2023 le premier sommet international sur les risques de l'IA. À Washington, on présentait le premier décret de grande envergure sur l'IA, l'*executive order* du président Joe Biden qui met en place 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 IA « sûre, sécurisée et digne de confiance ». Ce décret entend notamment élucider des problématiques liées à l'équité et aux droits civiques, en abordant également des applications spécifiques de l'IA.

À ces réglementations étatiques ou interétatiques s'ajoutent 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. Cette régulation en *patchwork* fait craindre une balkanisation de la gouvernance de l'IA, avec l'adoption fragmentée de normes concurrentes voire contradictoires, ce qui irait de fait à l'encontre de l'ambition d'une régulation universelle d'une technologie qui l'est tout autant.

## Les défis d'un consensus global

Derrière ces différences sur la régulation de l'IA se cachent en réalité des intérêts nationaux bien compris. Les États tentent de défendre leurs entreprises, qui estiment que les efforts de régulation vont à l'encontre de leur capacité à innover. Une tendance particulièrement visible dans les réticences françaises lors des tractations sur l'*AI Act* : derrière sa défense de l'innovation, Paris voulait assurer la 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.

Les grandes puissances s'efforcent de prendre le leadership en matière de gouvernance de l'IA, pour faire advenir des réglementations compatibles avec leurs ambitions nationales et pouvant, au passage, freiner leurs compétiteurs. À la course au développement technologique s'ajoute donc une vaste compétition entre les États, dont beaucoup rêvent de devenir la première puissance normative de l'IA. Le sénateur américain Chuck Shumer 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 » en matière d'IA.

Au-delà de cette recherche d'un avantage concurrentiel, les États veulent également imposer un modèle de valeurs au développement et au déploiement de l'IA, à l'image de l'UE qui insiste sur le respect des droits humains. Les pays en développement se concentrent quant à eux sur les conséquences sociales et les inégalités économiques créées par ces technologies émergentes. La régulation de l'IA devrait dès lors s'accompagner d'efforts systématiques pour améliorer l'accès à ces innovations : les économies en développement entendent en effet bénéficier de leur potentiel pour combler le fossé numérique, au lieu de l'aggraver.

Dans cette compétition normative internationale, l'UE jouit d'une certaine avance, puisqu'elle est arrivée au terme de ses travaux sur l'*AI Act*. Le décret de la Maison-Blanche et les réglementations chinoises montrent par ailleurs une volonté partagée d'endiguer les risques associés à ces innovations. Mais seule une approche véritablement universelle permettra de faire face aux risques qu'emporte le recours à l'IA.

Plusieurs initiatives vont déjà en ce sens. D'abord, le Partenariat mondial pour l'IA (PMIA), créé en 2019, réunit 29 membres (28 États et l'UE) dans le but de renforcer la coopération multi-acteurs 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 de l'homme et les libertés fondamentales.

L'Organisation de coopération et de développement économiques (OCDE) a, elle, élaboré des principes initialement adoptés en 2019 et mis à jour en mai 2024. Cette actualisation, qui vise à intégrer les nouveaux développements technologiques et politiques, témoigne de la nécessité d'une régulation *future proof* pour que les normes établies demeurent robustes et adaptées à leur objectif en dépit de l'évolution des technologies.

Les principes de l'OCDE bénéficient d'un rayonnement certain : les États membres se fondent en effet sur ces lignes directrices pour établir leurs édifices normatifs,

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.

Lors du G7 de mai 2023, sous présidence japonaise, a été lancé le Processus d'Hiroshima, qui doit permettre de définir les grands principes régissant l'usage de l'IA générative et plus largement des modèles d'IA avancés (*frontier AI*). Pilotée par les Japonais, cette initiative regroupe plusieurs groupes de travail créés pour aborder ces questions.

Au terme du premier *AI Safety Summit* à Londres, 28 États dont la Chine, les États-Unis et l'UE se sont par ailleurs accordés sur la 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. Enfin, l'Assemblée générale des Nations unies a adopté en mars 2024 un texte visant à établir des règles internationales pour les usages de l'IA.

La gouvernance de l'IA ne saurait être complète sans la participation active de ceux qui fabriquent ces technologies : les *Big Tech*. Comme le soulignait le Secrétaire général des Nations unies 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, du suivi et de l'atténuation des préjudices futurs, au regard du développement actuel de l'IA. » Ces entreprises ne sont pourtant pas toujours coopératives, parce qu'elles tendent à voir la régulation comme un obstacle à l'innovation. 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 internationale de l'IA.

La question est posée de l'opérationnalisation de ces grands principes. Bien qu'ils puissent guider les acteurs de l'IA dans leurs efforts pour développer une IA digne de confiance et fournir aux décideurs des recommandations pour des politiques efficaces, ils demeurent généraux, avec une large marge d'interprétation. Pour que les garde-fous soient robustes, il faut harmoniser les standards et traduire techniquement les principes sur lesquels les acteurs se sont accordés. Les organismes internationaux de standardisation, en particulier l'ISO, sont à ce titre cruciaux pour faire advenir une gouvernance de l'IA qui dépasse les vœux pieux et puisse véritablement contraindre les géants du numérique américains ou chinois.

L. de R.-R.

#### Pour en savoir plus

- M. Delmas-Marty, « Une boussole des possibles : gouvernance mondiale et humanismes juridiques », Paris, Collège de France, 2020.
- C. Kerrigan, *Artificial Intelligence: Law and Regulation*, Cheltenham Northampton, Edward Elgar Publishing, 2022.
- J. Tallberg, E. Erman, M. Furendal *et al.*, « The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research », *International Studies Review*, vol. 25, n° 3, septembre 2023.

