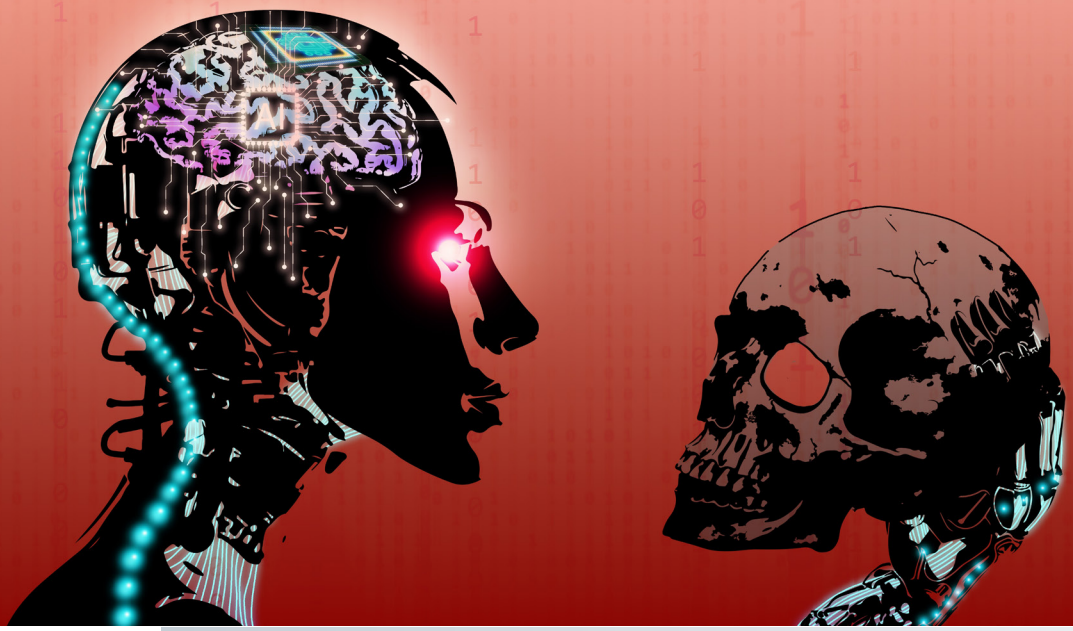
ÉTUDES
DE L'IFRI

CENTRE GÉOPOLITIQUE DES TECHNOLOGIES



JUIN
2024

Les risques de l'IA

Enjeux discursifs d'une technologie stratégique

Benjamin PAJOT

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 : 979-10-373-0878-8

Comment citer cette publication :
Benjamin Pajot, « Les risques de l'IA. Enjeux discursifs d'une technologie stratégique », *Études de l'Ifri*, Ifri, juin 2024.

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

Auteur

Benjamin Pajot est chercheur associé au Centre géopolitique des technologies de l'Ifri depuis février 2024 et chercheur indépendant en géopolitique numérique. Après avoir complété un master d'Histoire contemporaine à l'ENS-Lyon et un second en Études de guerre à l'Université Paris 1 Panthéon-Sorbonne, il a travaillé comme Chargé de mission sur les enjeux numériques et cyber au Centre d'analyse, de prospective et de stratégie du ministère de l'Europe et des Affaires étrangères (CAPS).

Ses recherches portent sur la rivalité technologique sino-américaine, les manipulations de l'information, l'impact sociopolitique des technologies émergentes et les biens communs numériques. Il a également travaillé sur les dimensions numériques et cyber du conflit russo-ukrainien.

Résumé

L'année 2023 aura été marquée par une ruée vers l'intelligence artificielle (IA) générative à tous niveaux – en particulier financier, médiatique et donc politique –, positionnant celle-ci au centre des discussions internationales comme peu de technologies auparavant. Si la vague d'innovations en cours justifie l'intérêt sans cesse renouvelé pour ce vaste amalgame de techniques, l'engouement et l'effroi qu'il suscite se fondent notamment sur des récits qu'il faut pouvoir décrypter. Ces récits tracent des horizons plus ou moins souhaitables selon les contextes. Entre quête d'une science sans limites, mythe de la singularité et risques d'extinction humaine, course à la puissance et peur du déclassement, nombreux sont les narratifs qui orientent aussi bien le développement de ces technologies que leur régulation balbutiante. Mettant aux prises une pluralité d'acteurs (leaders technologiques, États, organisations internationales, groupes de pression, organisations non gouvernementales, etc.), ils prennent le risque de faire naître des attentes disproportionnées et traduisent les profonds clivages de nos sociétés.

Cette étude se propose donc de retracer les principaux débats entourant l'IA aujourd'hui et d'en dégager les grandes lignes de force, afin d'identifier ce qui se joue derrière les discours et positionnements des uns et des autres. Les stratégies déployées sont multiples, allant de l'agitation de la menace géopolitique pour encourager les investissements au détournement de l'attention des régulateurs vers des risques à long terme, en passant par la condamnation ou la promotion de l'IA *open source* face à la concentration du marché. Il s'agit également d'analyser les risques communément associés au déploiement de ces technologies, qui sont au fondement des récits et des perceptions collectives. Ces risques menacent l'intégrité démocratique, l'environnement, la conduite de la guerre et la cybersécurité.

Cette étude suggère enfin que la bataille de la gouvernance à venir est déjà et continuera d'être profondément orientée par le rôle prescriptif de narratifs permettant de contrôler pour partie les débats politiques. Les États, tiraillés entre ces récits et leurs propres ambitions technologiques, rivalisent d'initiatives mais peinent à faire converger leurs visions. Dès lors, le risque serait de voir cette gouvernance internationale de l'IA réduite à une pure injonction collective et remise sans cesse au lendemain. La politique de sommets d'ampleur mondiale initiée notamment par le Royaume-Uni semble vouloir apporter un début de réponse, dont l'avenir permettra de déterminer si elle sera suivie d'effets.

Executive summary

The year 2023 has been marked by a rush towards generative artificial intelligence at all levels – particularly financial, media, and political – which has placed it at the center of international discussions like few technologies before. While the ongoing wave of innovations justifies the continually renewed interest in this vast amalgam of techniques, the enthusiasm and fear it arouses are based notably on narratives that must be deciphered. These narratives outline more or less desirable horizons depending on the contexts. Between the quest for limitless science, the myth of singularity, and the risks of human extinction, the race for power and fear of social, economic, and geopolitical decline, there are many narratives that guide both the development of these technologies and their nascent regulation. Engaging a plurality of actors (technological leaders, states, international organizations, pressure groups, NGOs, etc.), they risk creating disproportionate expectations and reflecting the deep divisions within our societies.

This study thus aims to trace the main debates surrounding AI today and to highlight the major forces at play in order to identify what is at stake behind the various discourses and positions. The strategies deployed are multi-faceted, ranging from stirring geopolitical threats to encourage investments and diverting regulators' attention to long-term risks to condemning or promoting open-source AI in the face of market concentration. The study also attempts to analyze the risks commonly associated with the deployment of these technologies, which underpin these narratives and collective perceptions. These risks threaten democratic integrity, the environment, warfare, and cybersecurity.

Finally, it suggests that the upcoming governance battle is already and will continue to be profoundly influenced by the prescriptive role of narratives that partly control political debates. States, torn between these narratives and their own technological ambitions, vie for initiatives but struggle to converge their visions. Consequently, the risk is that this international AI governance could be reduced to mere collective injunction, continually postponed. The policy of global summits initiated by the United Kingdom, in particular, seems to provide the beginnings of a response, but the future will tell whether it will be followed by action.

Sommaire

Introduction

De début mai à fin septembre 2023, une longue grève des scénaristes d'Hollywood a connu un retentissement inédit, en raison de son impact sur la production audiovisuelle mondiale. Aux revendications salariales qui constituaient le cœur de la mobilisation, s'ajoutaient des demandes de garanties face à la menace que l'intelligence artificielle (IA) faisait dorénavant peser sur ces métiers[1]. Loin de représenter une opportunité, l'IA était plutôt perçue comme la future remplaçante d'une corporation vouée à la disparition. En obtenant un encadrement de l'usage de ces technologies, cette grève a notamment réactualisé les débats sur les risques sociétaux engendrés par le recours à de tels outils, tandis que la rapide propagation d'hypertrucages vidéo (*deepfakes*) à travers le monde faisait monter les préoccupations d'ordre éthique et politique.

Depuis plus de deux ans, l'IA est ainsi au cœur de l'actualité internationale, portée par l'engouement populaire suscité par l'avènement de l'une de ses composantes : l'IA générative. Cette dernière désigne une vague d'innovations permise notamment par les progrès de l'apprentissage profond (*deep learning*) et du traitement automatique du langage naturel (*Natural Langage Processing* ou NLP) autorisant des systèmes informatiques à résoudre des tâches complexes et à « interpréter » le langage humain. Reposant sur l'exploitation statistique d'immenses jeux de données préalables, ces technologies ont rendu possible la génération « automatisée » de contenus textuels ou visuels complexes (code informatique, textes de loi, œuvres d'art) à partir d'instructions textuelles (*prompts*) relativement simples.

L'IA générative est aujourd'hui incarnée par les modèles de fondation (*foundation models)* généralistes et iconiques que sont GPT4 (développé par la société OpenAI), Gemini (Google) ou Llama (Meta). Elle a ensuite été popularisée par l'arrivée sur le marché de robots conversationnels comme ChatGPT (OpenAI), Bard (Google) ou Ernie Bot (Baidu), qui ont permis au grand public d'interagir directement avec des grands modèles de langage (*Large Language Model* ou LLM) en d'en éprouver les multiples capacités.

1. L'intelligence artificielle repose sur la création et l'application d'algorithmes et la mobilisation inférentielle de données la plupart du temps massives, à des fins d'imitation des fonctions cognitives (recherche, compréhension, génération de contenu, prise de décision). Elle désigne ainsi un répertoire de techniques très diverses (apprentissage statistique, apprentissage profond, apprentissage par transfert, réseaux antagonistes génératifs, etc.) pour des applications qui le sont tout autant (moteurs de recherche et algorithmes de classement et/ou de recommandation ; agents conversationnels ; reconnaissance et traduction ou génération d'image, de voix ou de texte ; prise de décision automatisée et systèmes autonomes ; etc.). Par souci de simplification, le terme générique d'IA, qui recouvre des réalités multiples, sera ici utilisé de manière uniforme.

Originellement développés par des *start-ups* ou des équipes de recherche relativement modestes, ces modèles sont aujourd'hui en majorité déployés par les grandes entreprises des secteurs technologiques américain et chinois, qui dominent l'industrie et la recherche en IA. Seuls acteurs à disposer des moyens à la fois humains et financiers considérables que requiert l'entraînement de tels modèles, les géants de la Silicon Valley (*Big Tech* – que sont notamment Google, Microsoft ou Amazon dans le paysage actuel de l'IA) sont parvenus à faire main basse sur ces technologies et à imposer une concentration rapide du marché à leur profit. À l'opposé de cette tendance monopolistique, le développement de modèles en source ouverte (*open source*) continue de représenter une alternative crédible. Celle-ci est soutenue à la fois par les communautés de recherche, les acteurs plus modestes de l'écosystème qui s'appuient sur ces dernières (BigScience, Stability AI, Hugging Face, Mistral, Aleph Alpha, Together AI ou encore EleutherAI) et par ceux des géants qui ont pris du retard dans la course ou ont intérêt à maintenir le marché ouvert (Meta, IBM, NVIDIA). L'*open source* fait dès lors l'objet de récits concurrents quant à sa sécurisation, dans l'espoir de renforcer sa légitimité, ou de la miner. Ces récits sont ensuite relayés dans une intense bataille à Washington, dans un contexte de rivalité accrue avec Pékin.

Face à la domination technologique sino-américaine, d'autres États tentent de se positionner dans ce qui ressemble à une folle course à l'IA, à l'image de la France, du Royaume-Uni, de l'Allemagne, l'Inde, du Canada, de la Corée du Sud ou encore des pays du Golfe. Ils déploient des stratégies propres, à l'appui de potentiels champions industriels nationaux. De son côté, l'Union européenne (UE) tente également de créer un nouvel « effet Bruxelles » (*Brussels Effect*) en forgeant l'une des premières régulations d'ampleur au travers de sa loi sur l'IA (*AI Act*), avec l'espoir de promouvoir ses normes et sa vision à l'échelle internationale[2]. Alors que chaque puissance tente de tirer son épingle du jeu et que la nécessité d'établir une gouvernance internationale de l'IA se fait de plus en plus pressante, une politique de sommets internationaux semble progressivement voir le jour, dans le sillage du sommet de Londres de novembre 2023.

Dans ce contexte de forte compétition aux niveaux tant politique et géopolitique qu'industriel, de multiples visions et discours s'opposent quant à l'appréciation des risques inhérents au développement de l'IA et des priorités en matière de gouvernance de ces technologies. En fonction des acteurs et des intérêts considérés, un primat à la régulation ou à l'innovation peut avoir cours. Des récits plus ou moins opportunistes sont ainsi construits

2. Le *Brussels Effect* désignait à l'origine l'influence qu'eut notamment l'adoption du régime général de protection des données personnelles (RGPD) sur le comportement des entreprises du secteur technologique et sur l'ensemble des politiques nationales de réglementation analogues à travers le monde. Par extension, il fait référence à la puissance normative européenne dans le champ numérique. Lire A. Bradford, *The Brussels Effect: How the European Union Rules the World*, Oxford, Oxford University Press, 2020.

et promus jusqu'au sein des appareils d'État et des organisations internationales par des personnalités en vue et des groupes de pression, afin de peser sur les grandes orientations politiques élaborées en leur sein. Ils opposent aujourd'hui deux courants principaux quant au futur de l'IA, tous deux issus de la Silicon Valley, l'un techno-enthousiaste voire solutionniste[3], l'autre pessimiste voire alarmiste.

Les divergences portent notamment sur la nature et la temporalité des risques ou des opportunités associés au déploiement de l'IA, allant de l'extinction de l'humanité jusqu'à la quête de l'immortalité. Ces positionnements ne sont toutefois pas orthogonaux et tendent parfois à se rejoindre en fonction des stratégies personnelles plus ou moins convergentes de leurs adeptes. En témoigne la quête largement partagée dans ces milieux de la création d'une « IA générale » (*Artificial General Intelligence* ou AGI), stade de développement à partir duquel l'IA deviendrait capable d'effectuer les mêmes tâches cognitives qu'un cerveau humain. L'avènement de l'AGI fascine ainsi diverses figures du secteur, à l'image de Sam Altman (à la tête d'OpenAI), Elon Musk (X, Tesla et Space X), Peter Thiel (Palanthir), Marc Zuckerberg (Meta), Demis Hassabis (DeepMind) ou Jensen Huang (NVIDIA). De fait, bon nombre de ces acteurs au rôle prescripteur ont avant tout pour but de poursuivre le développement du secteur et d'étendre le champ de leurs activités scientifiques et économiques.

Dès lors, l'enjeu demeure de faire la part des choses entre les risques réels et supposés, pour tenter de déconstruire ces récits et d'identifier non seulement les idéologies mais aussi les intérêts sous-jacents de ceux qui les portent dans l'espace public et dans les arcanes politiques. L'IA générative soulève à ce titre des inquiétudes légitimes quant à ses impacts sur les manipulations de l'information, la cybersécurité ou encore les biotechnologies. Face à la multiplicité des conséquences que le recours à l'IA pourrait engendrer en fonction des sociétés considérées, il s'agit également de prendre la mesure des risques les moins médiatiquement débattus, en particulier ceux qui concernent l'écologie ou l'accroissement des inégalités.

3. Le solutionnisme technologique désigne l'appétence voire la croyance de certains acteurs de la Silicon Valley en la possibilité de résoudre n'importe quel type de problème par la création ou le recours à une nouvelle technologie. Le concept a été forgé et popularisé par Evgeny Morozov dans *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist*, Londres, Allen Lane, 2013.

Petites et grandes promesses de l'intelligence artificielle

Moteur de transformation des sociétés ou agitation des imaginaires ?

Les progrès fulgurants de l'IA générative nous renvoient collectivement à des enjeux de perception divers. Convoquant des imaginaires partiellement nourris par la culture populaire (en particulier les ouvrages et films de science-fiction) et renforcés par des discours et narratifs produits à dessein par les acteurs du secteur, l'IA est tour à tour perçue comme promesse de progrès et d'émancipation pour le plus grand nombre, ou comme un enjeu de sécurité nationale majeur – voire une menace existentielle.

Ce type de rapport ambivalent à la technologie n'est pas nouveau en soi et s'inscrit dans une forme de continuité historique : les inventions de l'électricité, du moteur à explosion et de l'informatique ont déclenché des fractures similaires dans le débat public. Néanmoins, il semble qu'un tournant discursif entérinant le rôle sociétal central de l'IA et de la gouvernance algorithmique se soit opéré[4], aussi bien dans la communication des entreprises que dans celle des institutions politiques. Si la technologie s'est régulièrement vue conviée à résoudre des problèmes qui étaient loin d'être uniquement techniques, l'IA est aujourd'hui brandie de Washington à Pékin comme la solution potentielle aux grands problèmes de société contemporains[5] (climat, inégalités, croissance et productivité en baisse, pandémies, etc.). Ceci contribue à conférer à ces technologies une légitimité intrinsèque et assoit le caractère inéluctable de leur développement.

Ceci tient notamment au fait que les domaines d'application sont supposément multiples : à titre d'exemples divers, l'IA pourrait ainsi renforcer la fiabilité des prévisions météorologiques et la prévention des catastrophes naturelles[6], le suivi de la santé mentale[7], l'efficacité du recyclage[8] ou encore la

4. C. Katzenbach, « "AI Will Fix This" – The Technical, Discursive, and Political Turn to AI in Governing Communication », *Big Data & Society*, vol. 8, n° 2, octobre 2021.
5. M.-E. Bobillier-Chaumon, « L'IA n'est pas considérée comme une solution parmi d'autres, mais comme la solution à tous les problèmes de l'organisation du travail », *Le Monde*, 5 avril 2024, disponible sur : www.lemonde.fr.
6. « L'IA pour lutter contre le changement climatique et favoriser la durabilité environnementale », Inria, 6 juillet 2023, disponible sur : www.inria.fr.
7. S. Cabut et P. Santi, « L'intelligence artificielle au secours du suivi de la santé mentale », *Le Monde*, 16 février 2024, disponible sur : www.lemonde.fr.
8. N. Rivero, « How the World of Recycling Is About to Be Transformed », *The Washington Post*, 7 février 2024, disponible sur : www.washingtonpost.com.

traçabilité des navires[9]. En facilitant la découverte de nouvelles molécules et nouveaux matériaux[10], elle permet d'envisager de potentiels applicatifs et gains d'efficience dans des domaines aussi variés que la médecine, la chimie, le bâtiment, la logistique ou le stockage énergétique. Elle offre ainsi la perspective d'un avenir désirable fait de nouvelles découvertes scientifiques majeures (déchiffrage de traces archéologiques, compréhension de la biodiversité, exploration spatiale, etc.)[11]. L'ensemble de ces promesses s'accompagne de récits d'un « nouvel âge des Lumières[12] », d'une « ère d'abondance[13] », d'une inéluctabilité du progrès technologique à même de transformer la science dans son ensemble[14] et, partant, le futur de l'humanité.

Dans ce contexte, l'IA générative est régulièrement convoquée comme la solution potentielle à des problèmes de nature hautement complexe, à l'image de la conception de microprocesseurs[15], sans qu'il soit toujours possible de distinguer ce qui relève de l'argument marketing ou du réel saut technologique. À ce titre, elle est aussi opportunément présentée par les plateformes de réseaux sociaux comme la panacée face aux enjeux de gestion et de modération des contenus en ligne, évacuant le soubassement sociopolitique de ces derniers. Or les faiblesses des modèles restent nombreuses, sur ce terrain comme sur d'autres ; malgré les réalités sur lesquelles se fonde l'engouement général, il faut rappeler que ces technologies connaissent encore de grandes difficultés à réaliser des tâches multiples, voire à résoudre certains problèmes relativement élémentaires[16].

9. « Vast Amounts of the World's Shipping Sails Unseen », *The Economist*, 3 janvier 2024, disponible sur : www.economist.com.
10. J. Kim, « Google Deepmind's New AI Tool Helped Create More Than 700 New Materials », *MIT Technology Review*, 29 novembre 2023, disponible sur : www.technologyreview.com.
11. J. Marchant, « AI Reads Text from Ancient Herculaneum Scroll for the First Time », *Nature*, 12 octobre 2023, disponible sur : www.nature.com ; L. Parshley, « Artificial Intelligence Could Finally Let Us Talk with Animals », *Scientific American*, 1er octobre 2023, disponible sur : www.scientificamerican.com.
12. Ce sont notamment les termes de Yann Le Cun, considéré comme l'une des figures les plus importantes de la recherche en IA. Sundar Pichai (P.-D.G. de Google) déclarait quant à lui dès 2018 que l'IA était « plus profonde que le feu ou l'électricité ». Lire C. Clifford, « Google CEO: A.I. Is More Important Than Fire or Electricity », CNBC, 1er février 2018, disponible sur : www.cnbc.com.
13. A. Acharya, « How AI Will Usher in an Era of Abundance », *a16z* (blog d'Andreessen Horowitz), 7 février 2024, disponible sur : a16z.com.
14. Notamment en favorisant l'émergence de « robots laborantins », la vérification et la reproductibilité des expériences, ou en identifiant des recoupements fructueux au sein des communautés scientifiques et de l'abondante littérature déjà disponible. « Could AI Transform Science Itself? », *The Economist*, 13 septembre 2023, disponible sur : www.economist.com.
15. B. Lin « Designing Chips Is Getting Harder: These Engineers Say Chatbots and AI Can Help », *The Wall Street Journal*, 6 février 2024, disponible sur : www.wsj.com.
16. Sans s'attarder sur les biais et hallucinations des modèles, ou sur le rapport de l'IA à la créativité qui animent les débats dans nombre de milieux, ces technologies buttent encore jusqu'ici sur la hiérarchisation de l'information, la compréhension du contexte, la récursivité, l'explicabilité et la fiabilité des résultats dans le temps, pour ne citer que quelques exemples.

Une surenchère rhétorique et marketing

Qu'elle soit louée ou décriée, l'IA générative fait donc l'objet d'une inflation discursive qui dissimule diverses stratégies. Dans le sillage de ses progrès, de vieux réflexes pavloviens semblent avoir ressurgi : les discours techno-optimistes et solutionnistes se succèdent, en particulier chez des grandes figures du secteur, flirtant régulièrement avec le messianisme, le transhumanisme, l'utopie voire l'eugénisme[17]. Ces acteurs sont également désignés ou s'identifient comme des « accélérationnistes efficaces » *(effective accelerationists*), par opposition aux « altruistes efficaces » ou pessimistes (*effective altruists* ou *doomers*[18]), dont les discours alarmistes[19] – pour ne pas dire apocalyptiques et messianiques – sont eux aussi récurrents.

Paradoxalement, ces récits en apparence contraires peuvent être le fait des mêmes personnes, chez qui le mythe de la singularité – stade à partir duquel l'IA dépasserait l'ensemble des capacités humaines et pourrait ainsi représenter une menace – constitue une toile de fond commune. Un nouveau terme a ainsi été forgé pour désigner tout à la fois le chevauchement de ces diverses idéologies, leur perméabilité et leur dangerosité : le « tescrealisme[20] » (*tescrealism*). Au-delà, la sincérité des positionnements peut être questionnée, dans la mesure où ces discours, sous couvert de préoccupations éthiques, peuvent en réalité avoir pour but de ralentir des concurrents, en renforçant les barrières à l'entrée ou en réclamant des moratoires sur le développement d'IA plus avancés[21].

17. A. Piquard, « Derrière l'intelligence artificielle, le retour d'utopies technologiques », *Le Monde*, 13 juin 2023, disponible sur : www.lemonde.fr.

18. L'*effective altruism* désigne un courant issu de la Silicon Valley et très influent à Washington. Initialement centré sur la transformation positive des sociétés grâce à l'IA, il semble à présent davantage préoccupé, sous couvert de « long-termisme », par l'avènement d'une prétendue IA générale et par les risques existentiels que celle-ci ferait peser sur l'espèce humaine. Son poids au sein des administrations et cercles de réflexion américains contribue à orienter les débats en ce sens, aux dépens de préoccupations plus immédiates. Lire B. Bordelon, « When Silicon Valley's AI Warriors Came to Washington », *Politico*, 30 décembre 2023, disponible sur : www.politico.com.

19. K. Roose, « A.I. Poses "Risk of Extinction", Industry Leaders Warn », *The New York Times*, 30 mai 2023, disponible sur : www.nytimes.com.

20. Le terme résulte de l'acronyme de transhumanisme (*transhumanism*, soit l'augmentation des capacités physiques et mentales de l'homme par la science et notamment par la technologie), extropisme (*extropianism*, un sous-courant du transhumanisme rejetant l'entropie des systèmes physiques pour leur préférer une croissance illimitée grâce à la science), « singularitarisme » (*singularitarianism*, croyance en l'avènement de la singularité technologique), cosmisme (*cosmism*, courant d'origine russe qui prône la recherche de l'immortalité en symbiose avec le « cosmos », ainsi que la conquête spatiale), rationalisme (*rationalism*, mouvement qui prône l'optimisation des capacités intellectuelles et de la rationalité aux échelles individuelle et collective) et « altruisme efficace ». Le « tescrealisme » identifie et dénonce le socle commun de ces diverses idéologies, à savoir l'avènement d'une prétendue IA générale. Lire E. Torres, « The Acronym Behind Our Wildest AI Dreams and Nightmares », *Truthdig*, 15 juin 2023, disponible sur : www.truthdig.com.

21. Cela peut se traduire par des volontés d'élever les prérequis en matière de redevabilité (déclarations préalables au développement de modèles d'IA avancés, établissement de régimes de licences, restrictions sur les modèles *open source*, « pause » complète sur les expériences généralistes, etc.). Dans le même ordre d'idée, des lettres ouvertes ont dénoncé à plusieurs reprises les risques existentiels qu'engendrerait l'IA générative ; la plus retentissante, lancée par l'organisme de lobbying Futur of Life Institute, exigeait

Surtout, ils contribuent opportunément à détourner l'attention d'autres enjeux plus immédiats (régulation, fiscalité, multiplicité des impacts sociaux et sociétaux[22]). Les entreprises et investisseurs américains déploient ainsi des stratégies agressives pour tenter de peser sur les orientations politiques à la Maison-Blanche et au sein de l'administration américaine[23], au Congrès mais aussi en Europe – notamment pour focaliser la régulation sur les risques à long terme plutôt qu'à court terme, afin de se laisser davantage de champ d'action[24]. On ne peut dès lors que souligner l'absence de cohérence chez ceux qui, à l'instar d'Elon Musk ou de Sam Altman, tiennent des propos alarmistes tout en s'engageant dans une course à l'AGI – qu'ils disent craindre et œuvrent pour limiter une régulation qu'ils font mine d'appeler de leurs vœux –, se déchargeant ainsi sur le régulateur de leur propre responsabilité. D'une année à l'autre, le discours ambiant semble néanmoins subir leur influence et évoluer au gré de leurs revirements, penchant tour à tour du côté techno-optimiste ou du côté alarmiste[25].

Il faut cependant faire la part commerciale de cette inflation discursive. À l'instar des déclarations claironnantes qui sont monnaie courante dans l'informatique quantique, chaque entreprise clame régulièrement avoir développé le dernier LLM le plus puissant : les sorties de modèles dont les appellations pompeuses singent la puissance se succèdent sans discontinuer[26], sans que les méthodes d'évaluation de leurs performances fassent l'unanimité[27]. Plus encore, le développement l'AGI devient l'horizon attendu : surjouant la rivalité commerciale et la course à l'innovation, les acteurs majeurs du secteur (OpenAI, Google, Meta, Anthropic) ont tous affirmé qu'il

notamment un moratoire sur le développement des modèles et sur les recherches. Elon Musk, soutien financier de cet organisme, figurait parmi les signataires de la lettre, alors qu'il est lui-même à la tête d'entreprises engagées dans la course à l'IA et concurrentes des principaux leaders.

22. Ce d'autant plus que ce prétendu « risque d'extinction » fait l'objet d'un lobbying intense par des organismes (Open Philanthropy, Future of Life Institute, Center for AI Policy, Center for AI Safety, Foundation for American Innovation) financés par des grandes figures du capitalisme numérique ayant investi massivement dans l'IA. Ce lobbying porte jusqu'en Europe puisqu'Ursula von der Leyen en a repris les grandes lignes dans son discours de l'Union de 2023. Lire B. Bordelon, « AI Doomsayers Funded by Billionaires Ramp Up Lobbying », *Politico*, 23 février 2024, disponible sur : www.politico.com.

23. L'influence des *effective altruists* est perceptible jusque dans les agences fédérales – comme en témoigne par exemple la nomination du « pessimiste » Paul Christiano (ancien d'OpenAI) au sein du US AI Safety Institute – et transparaît ainsi dans des études « sérieuses », à l'image de celle commandée récemment par le Département d'État américain, qui adhère au récit du risque d'extinction de l'espèce humaine par l'IA. Lire A. Belanger, « Feds Appoint "AI Doomer" to Run AI Safety at US Institute », *Ars Technica*, 17 avril 2024, disponible sur : https://arstechnica.com ; M. Egan, « AI Could Pose "Extinction-level" Threat to Humans and the US Must Intervene, State Dept.-Commissioned Report Warns », CNN, disponible sur : www.edition.cnn.com.

24. Un réseau de collaborateurs parlementaires au Congrès aurait ainsi été financé par l'Open Philanthropy, organisation à but non lucratif proche des entreprises du secteur. Lire B. Bordelon, « How a Billionaire-backed Network of AI Advisers Took over Washington », *Politico*, 13 octobre 2023, disponible sur : www.politico.com.

25. C'est notamment sensible dans les revirements au forum de Davos. Lire C. Zakrzewski, « The Davos Elite Embraced AI in 2023. Now They Fear It », *The Washington Post*, 18 janvier 2024, disponible sur : www.washingtonpost.com.

26. Ainsi de l'« Olympus » d'Amazon ou du « Gemini Ultra » de Google.

27. W. D. Heaven, « AI Hype Is Built on High Test Scores. Those Tests Are Flawed », *MIT Technology Review*, 30 août 2023, disponible sur : www.technologyreview.com.

s'agissait d'un objectif de long terme[28]. D'autres vantent l'imminence de l'émergence d'une AGI, en fonction de la fluctuation des définitions et de ce qu'impliquerait un tel seuil de développement[29]. Associé au mythe de la singularité, cet objectif contribue là encore à façonner les perceptions et à détourner opportunément l'attention des régulateurs pour mieux servir les intérêts des leaders du secteur[30].

Au-delà de cette question, la concurrence des récits se noue aussi autour de l'opposition consacrée entre régulation et innovation. Celle-ci est notamment au cœur des enjeux de propriété intellectuelle, profondément malmenée par les pratiques des grands acteurs du secteur[31] – lesquels ne défendent les droits d'auteur que pour mieux freiner le développement de concurrents dont la taille plus modeste ne leur permet pas de concevoir des bases de données libres de droits. Les manœuvres de ces derniers pour peser sur les discours et la législation en matière de droit de la propriété intellectuelle et de droits d'auteur sont régulières[32], de même que leurs tentatives d'accès à des données « propriétaires[33] ».

Ce clivage se mue en véritable lutte, alors que les premières plaintes pour non-respect du droit d'auteur ont été déposées par des acteurs majeurs[34], que la grève des scénaristes d'Hollywood a permis d'obtenir des garanties de non-recours à l'IA[35], et que des tentatives de protéger les œuvres

28. A. Heath, « Mark Zuckerberg's New Goal Is Creating Artificial General Intelligence », *The Verge*, 18 janvier 2024, disponible sur : www.theverge.com.
29. C'est notamment le cas de Jensen Huang, P.-D.G. de NVIDIA, qui a déclaré que l'IA générale (dont il a préalablement revu les critères à la baisse) pourrait voir le jour dans les cinq ans. Lire S. Nellis, « Nvidia CEO Says AI Could Pass Human Tests in Five Years », Reuters, 2 mars 2024, disponible sur : www.reuters.com. Une équipe de recherche de Microsoft estime également que Chat GPT4 montrerait déjà des « étincelles d'IA générale », tandis qu'une équipe de Google tente de redéfinir les termes du débat pour mieux l'influencer. Lire F. Coll. « Sparks of Artificial General Intelligence: Early Experiments with GPT-4 », *Arxiv*, 22 mars 2023, disponible sur : https://arxiv.org.
30. M. O'Shaughnessy, « How Hype Over AI Superintelligence Could Lead Policy Astray », Carnegie Endowment for International Peace, 14 septembre 2023, disponible sur : https://carnegieendowment.org.
31. Mus par l'habitus de l'innovation sans permission (*permissionless innovation*) profondément inscrit dans la culture de la Silicon Valley, les acteurs du secteur se plaisent à affirmer qu'il leur est impossible de développer leurs modèles sans recourir à des contenus protégés par des droits. Lire D. Milmo, « "Impossible" to Create AI Tools Like ChatGPT Without Copyrighted Material, OpenAI Says », *The Guardian*, 8 janvier 2024, disponible sur : www.theguardian.com.
32. À l'image de la tentative d'OpenAI d'influencer la législation du secteur par le Congrès américain. Lire B. Bordelon, « The Fingerprints on a Letter to Congress About AI », *Politico*, 23 octobre 2023, disponible sur : www.politco.com.
33. K. Paul et A. Tong, « Inside Big Tech's Underground Race to Buy AI Training Data », Reuters, 5 avril 2024, disponible sur : www.reuters.com.
34. À l'instar du *New York Times* qui attaque Microsoft et OpenAI en justice, ou des plaintes d'auteurs contre OpenAI ou encore Nvidia. Lire M. M. Grynbaum et R. Mac, « The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work », *The New York Times*, 27 décembre 2023, disponible sur : www.nytimes.com ; M. Zahn, « Authors Lawsuit Against OpenAI Could "Fundamentally Reshape" Artificial Intelligence, According to Experts », ABC News, 25 septembre 2023, disponible sur https://abcnews.go.com ; A. Belanger, « Nvidia Sued over AI Training Data as Copyright Clashes Continue », *Ars Technica*, 11 mars 2024, disponible sur : https://arstechnica.com.
35. C. Melas et D. Romero, « Writers Strike Negotiations Hung Up on Language over AI, Sources Say », NBC News, 24 septembre 2023, disponible sur : www.nbcnews.com.

des artistes voient le jour – y compris en ayant recours au *data poisoning*[36]. En sus de cette question d'ordre social, économique et juridique, l'opposition entre régulation et innovation prend également des atours stratégiques. Car le narratif de la course technologique se trouve ici poussé à son paroxysme pour défendre une sous-régulation du secteur, laquelle serait la seule garantie de préserver la capacité d'innovation et l'avance prise sur les concurrents étrangers.

Or, il faut rappeler que le mantra selon lequel la régulation tue l'innovation (et mettrait ainsi en danger les États occidentaux sur le plan géopolitique) est une antienne dont les fondements ne sont pas suffisamment établis[37]. Ce discours, porté notamment par les grands acteurs américains du secteur, se double généralement d'un narratif sur le risque de dépassement technologique par la Chine, qui, dans le domaine de l'IA générative, reste encore à démontrer[38].

À rebours de cette posture, seule l'UE cherche aujourd'hui à déployer un récit alternatif tendant à démontrer que régulation et innovation sont ne sont pas antinomiques mais complémentaires, et que la réglementation permet au contraire d'offrir un cadre juridique clair dans lequel développer sereinement ces technologies. Ce récit est notamment au fondement des tractations ayant abouti à la loi européenne sur l'IA (*AI Act*), mais trouve jusqu'ici peu d'échos dans les pays du Sud. Ces derniers perçoivent les tentatives de régulation à la fois comme une nécessité pour corriger les externalités négatives qu'ils sont les premiers à subir, mais aussi comme un obstacle potentiel à leur émancipation par la technologie. Confrontés au risque de se voir réduits à un rôle de simple maillon inférieur de la chaîne d'approvisionnement et de laissés-pour-compte face à un développement de l'IA qui nécessite de nombreuses ressources, ces pays connaissent par conséquent des débats moins vifs qu'ailleurs. L'Inde semble ainsi poursuivre un équilibre précaire entre opportunités et risques, qui souffre encore des lacunes de sa stratégie en matière de réglementation[39].

36. M. Heikkilä, « This New Data Poisoning Tool Lets Artists Fight Back Against Generative AI », *MIT Technology Review*, 23 octobre 2023, disponible sur www.technologyreview.com.
37. A. Bradford, *Digital Empires: The Global Battle to Regulate Technologies*, Oxford, Oxford University Press, 2023.
38. Sam Altman a notamment considéré lors de son audition de mai 2023 au Sénat qu'une régulation trop importante pourrait faciliter la montée en puissance de la Chine ou d'autres rivaux des États-Unis. La surestimation potentielle des capacités chinoises créerait ici un paravent utile pour amoindrir des efforts de régulation nécessaires. Lire H. Toner, J. Xiao et J. Ding, « The Illusion of China's AI Prowess », *Foreign Affairs*, 2 juin 2023, disponible sur www.foreignaffairs.com.
39. A. Mohanty et S. Sahu, « India's AI Strategy: Balancing Risk and Opportunity », Carnegie Endowment for International Peace, 22 février 2024, disponible sur : https://carnegieendowment.org.

Quand l'« AI for Good » s'éloigne

Alors que les impacts sociaux-économiques du secteur commencent à se faire sentir et que des pratiques controversées (reconnaissance faciale, surveillance algorithmique, contournement de la propriété intellectuelle, etc.) échappent encore dans une large mesure à la régulation[40], les efforts de ses leaders pour développer des IA tournées vers la promotion de l'intérêt général et du bien commun restent pour le moins fragiles. En dépit des déclarations d'intention, la volonté réelle des principaux acteurs peut ainsi être questionnée : à mesure que les cas d'usages se précisent, les équipes de Meta, Google, Microsoft et Amazon dédiées aux aspects « éthiques » ou aux « usages responsables » de l'IA se dépeuplent au profit des équipes de développement[41]. Qui plus est, la notion même « d'IA au service de l'humanité » (« AI for Good ») se voit instrumentalisée par la concurrence féroce à laquelle se livrent les entreprises[42].

Si dans sa phase de développement initial l'IA générative – selon les engagements initiaux des entreprises et du fait de la vigilance des employés du secteur[43] – ne devait pas servir des fins militaires, cette préoccupation semble aujourd'hui être passée au second plan. Malgré ses promesses initiales, OpenAI s'est ainsi rapproché du Pentagone pour lui fournir des outils de cybersécurité[44], quand Palantir ou Clearview AI ont mis leurs technologies à disposition non seulement du gouvernement américain mais aussi des autorités ukrainiennes. Les opérations américaines au Moyen-Orient et israéliennes à Gaza entérinent de fait le recours problématique à

40. Certains considèrent ainsi que ces entreprises sont en train de contracter une « dette éthique » à l'égard de nos sociétés. Lire C. Fiesler, « AI Has Social Consequences, But Who Pays the Price ? Tech Companies' Problem with Ethical Debt », *The Conversation,* 19 avril 2023, disponible sur : https://theconversation.com.
41. 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.
42. Elon Musk semble ainsi vouloir intenter un procès contre Open AI et son P.-D.G. pour « violation de ses principes fondateurs », accusant l'entreprise de privilégier le profit à l'intérêt général, contrairement à ses engagements initiaux. Lire A. Satariano, C. Metz et T. Mickle « Elon Musk Sues OpenAI and Sam Altman for Violating the Company's Principles », *The New York Times*, 1er mars 2024, disponible sur : www.nytimes.com.
43. Voir par exemple cette lettre ouverte publiée par le Futur of Life Institute, dans laquelle les divers acteurs du secteur s'engageaient en 2018 à ne pas travailler sur le développement d'armes autonomes. Lire « Lethal Autonomous Weapons Pledge », Future of Life Institute, 6 juin 2018, disponible sur : https://futureoflife.org. Depuis, OpenAI a notamment mis à jour ses conditions d'utilisation, ouvrant potentiellement la voie à des usages civilo-militaires de ChatGPT. Lire S. Biddle, « OpenAI Quietly Deletes Ban on Using ChatGPT for "Military and Warfare" », *The Intercept*, 12 janvier 2024, disponible sur : https://theintercept.com. Par ailleurs, la mobilisation des employés de Google avait par exemple eu raison du contrat de l'entreprise avec le projet Maven (plateforme d'interconnexion de données militaires reposant sur diverses technologies d'IA) auquel Amazon, Microsoft et d'autres sont restés parties. Lire N. Statt, « Google Reportedly Leaving Project Maven Military AI Program After 2019 », *The Verge*, 1er juin 2018, disponible sur : www.theverge.com. Pour autant Google est loin d'avoir coupé tout lien avec le Pentagone ensuite. Lire T. Simonite, « 3 Years After the Project Maven Uproar, Google Cozies to the Pentagon », Wired, 18 novembre 2021, disponible sur : www.wired.com.
44. B. Stone et M. Bergen, « OpenAI Is Working with US Military on Cybersecurity Tools », *Bloomberg,* 16 janvier 2024, disponible sur : www.bloomberg.com.

ces outils[45], tandis que l'Ukraine est devenue le laboratoire à ciel ouvert des usages militaires de l'IA – les nécessités du champ de bataille autorisant des expérimentations qui n'auraient pu être conduites autrement[46]. Dans le même temps, l'augmentation exponentielle du nombre de capteurs sur le champ de bataille implique de développer des capacités d'IA pour traiter les remontées d'information ne pouvant plus être analysées par des opérateurs humains. Tout ceci ouvre la voie à des usages militaires de l'IA plus systématiques, à mesure que le secteur privé acquiert expérience et réputation sur le terrain, et que les appareils de défense se dotent de stratégies d'emploi de plus en plus affirmées[47]. Tandis que des expériences alarmantes pointent l'acuité de tels enjeux[48], les craintes anciennes d'un avènement prochain des systèmes d'armes létales autonomes (SALA) ressurgissent avec d'autant plus d'intensité qu'une régulation internationale de ces armements semble aujourd'hui illusoire[49].

Vers un énième « hiver de l'IA » ?

Sans nier l'évolution majeure que représente l'avènement de l'IA générative, il faut néanmoins garder à l'esprit le caractère cyclique de l'histoire de l'IA, et le fait que chaque vague d'innovation successive a ensuite été suivie d'« hivers » de stagnation au cours desquels les progrès n'ont pas été à la hauteur des annonces et des attentes[50].

45. K. Manson, « AI Warfare Is Already Here », *Bloomberg*, 28 février 2024, disponible sur : www.bloomberg.com ; H. Davies, B. McKernan et D. Sabbagh, « "The Gospel": How Israel Uses AI to Select Bombing Targets in Gaza », *The Guardian*, 1er décembre 2023, disponible sur : www.theguardian.com ; S. Frankel, « Israel Deploys Expansive Facial Recognition Program in Gaza », *The New York Times*, 27 mars 2024, disponible sur : www.nytimes.com ; Y. Abraham « "Lavender": The AI Machine Directing Israel's Bombing Spree in Gaza », *+972 Magazine*, 3 avril 2024, disponible sur : www.972mag.com.

46. L'IA contribue directement à divers objectifs sur le terrain : déploiement de systèmes d'armes autonomes, observation et reconnaissance, identification et classification des cibles, analyse et prédiction des menaces, logistique, cybersécurité, guerre électronique, diagnostic médical et prise en charge des blessés, etc. Lire V. Bergengruen, « How Tech Giants Turned Ukraine Into an AI War Lab », *Time*, 8 février 2024, disponible sur : https//time.com.

47. « Data, Analytics, and Artificial Intelligence Adoption Strategy », Department de la Défense américain, 27 juin 2023 ; F. Bajak, « Pentagon's AI Initiatives Accelerate Hard Decisions on Lethal Autonomous Weapons », Associated Press, 25 novembre 2023, disponible sur : https://apnews.com. L'armée américaine déploierait aujourd'hui près de 800 projets d'IA aux objectifs variés, allant de la modélisation des risques, en passant par le traitement des données des capteurs d'armes, la planification des itinéraires de réapprovisionnement en munitions, jusqu'à l'aide à la détection et destruction de cibles ennemies. La France n'est pas en reste et a notamment créé une enveloppe de 300 millions d'euros pour développer l'IA militaire. Lire A. Bauer, « IA : la France dévoile son plan pour devenir la première puissance militaire d'Europe », *Les Échos*, 8 mars 2024, disponible sur : www.lesechos.fr.

48. « US Air Force Denies Running Simulation in Which AI Drone 'Killed' Operator », *The Guardian*, 2 juin 2023, disponible sur : www.theguardian.com.

49. L. de Roucy-Rochegonde, *La Guerre à l'ère de l'intelligence artificielle. Quand les machines prennent les armes*, Paris, PUF, à paraître.

50. D. Cardon, J.-P. Cointet et A. Mazières. « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle », *Réseaux*, vol. 211, n° 5, 2018, p. 173-220.

Alors que la tendance actuelle à résumer de plus en plus la diversité de l'IA au seul *machine learning* – et au-delà, à l'AGI – expose aussi bien à des déconvenues qu'à minorer d'autres champs de recherche (IA causale, neurosymbolique, vision par ordinateur, *small data*, etc.[51]), le marché, qui dicte pour partie le rythme actuel de progression de l'IA générative, pourrait finir par se détourner de cette dernière en cas de premiers retours sur investissements insuffisants et/ou d'espoirs déçus[52]. Certains craignent dès lors que le battage médiatique et la course folle aux investissements ne créent une bulle semblable à celle que le secteur des cryptomonnaies a connue il y a peu, et n'occulte paradoxalement les progrès réels accomplis[53]. Dans la mesure où bon nombre de logiciels passés de mode et inutilisés encombrent déjà les parcs informatiques, certains logiciels d'IA pourraient vite connaître un sort similaire, passés l'engouement initial et la surenchère commerciale qui l'accompagne. Il ne peut donc être totalement exclu que l'IA générative, confrontée à la réalité de ses limites techniques comme physiques, finisse par plonger à son tour dans un nouvel « hiver », comme le soutiennent certains techno-sceptiques[54]. Traversée par divers récits qui tendent à en proposer un miroir déformant, l'IA doit donc faire l'objet d'une analyse dépassionnée pour évaluer ses impacts substantiels.

51. L'IA causale (*causal AI*) est une approche dans laquelle le modèle effectue des déductions en établissant des chaînes de causalité plutôt que des corrélations entre diverses occurrences, ce qui permet notamment une plus grande explicabilité et traçabilité des résultats obtenus. L'IA neurosymbolique (*neurosymbolic AI*) concilie les deux courants historiques de modélisation (symboliste et connexionniste) pour tenter de modéliser des systèmes cognitifs plus complets. La vision par ordinateur (*computer vision*) consiste à doter les modèles d'IA de capacités de perception, d'interprétation et d'apprentissage par les images et vidéos analogues à celle du système visuel humain. Par opposition au *Big Data*, l'approche du *small data* cherche à développer des modèles avec des ensembles de données d'entraînement beaucoup plus restreints mais plus qualitatifs, ouvrant des perspectives en matière de réduction d'impact carbone et de barrières à l'entrée pour le développement de modèles complexes.

52. A. Piquard, « "La hype" autour de l'intelligence artificielle risque de créer des déceptions », *Le Monde*, 18 avril 2024, disponible sur : www.lemonde.fr.

53. J. Thornhill, « Huge AI Funding Leads to Hype and "Grifting", Warns DeepMind's Demis Hassabis », *Financial Times*, 31 mars 2024, disponible sur www.ft.com ; R. Foroohar, « Beware AI Euphoria », *Financial Times*, 24 mars 2024, disponible sur : www.ft.com.

54. V. J. Carchidi, « Are We Due for an AI Winter? », *The National Interest*, 3 septembre 2023, disponible sur : https://nationalinterest.org.

Entre risques réels et fantasmés

Désinformation et manipulations de l'information : changement d'échelle

Aujourd'hui bien documentées, les fameuses « hallucinations » des *chatbots* – qui les conduisent notamment à déformer voire inventer des informations de toutes pièces – demeurent un sujet de préoccupation[55]. Mais c'est avant tout la production à l'aide d'IA générative de contenus faux ou sciemment manipulés qui constitue le risque de désinformation le plus sérieux. Parce qu'il facilite la rédaction et la mise en ligne d'articles, le recours à l'IA générative induit une inflation incontrôlable et incontrôlée de la production de contenus potentiellement problématiques[56]. Les techniques d'entraînement des LLM sont également en jeu : des *chatbots* pernicieux, conçus à des fins de test des systèmes de détection et de modération automatisés, ont ainsi réussi à mettre en échec ces systèmes[57] – rappelant au passage la rapide obsolescence de ces outils et les limites du recours à l'IA pour résoudre les problèmes créés par elle-même[58].

L'IA générative fait alors peser un risque non négligeable sur l'intégrité des espaces démocratiques en ligne et hors ligne. Le seuil de tolérance juridique aux manipulations à caractère politique semble progressivement s'abaisser[59], mais nombre d'acteurs politiques refusent encore de se fixer des limites dans le recours à ces outils et contribuent directement à une logique

55. C. Metz, « Chatbots May "Hallucinate" More Often Than Many Realize », *The New York Times*, 6 novembre 2023, disponible sur : www.nytimes.com.
56. À tel point que Google serait par exemple en proie à des difficultés pour maintenir les performances et la pertinence de son moteur de recherche. Lire J. Koebler, « Google Search Really Has Gotten Worse, Researchers Find », *404 Media*, 18 janvier 2024, disponible sur : www.404media.co ; J. Cox, « Google News Is Boosting Garbage AI-Generated Articles », *404 Media*, 16 janvier 2024, disponible sur : www.404media.co.
57. D. Rafieyan, « Researchers at Anthropic Taught AI Chat Bots How to Lie, and They Were Way Too Good at It », *Business Insider*, 25 janvier 2024, disponible sur : www.businessinsider.com.
58. N. Jones, « How Journals Are Fighting Back Against a Wave of Questionable Images », *Nature*, 12 février 2024, disponible sur : www.nature.com.
59. Comme en témoigne la procédure engagée par le New Hampshire contre l'entreprise Life Corporation qui aurait effectué des appels téléphoniques ayant pour but de décourager les électeurs démocrates de voter lors des primaires de l'État, le tout en recourant à une IA reproduisant la voix du président Biden. Lire C. Zakrewski et P. Verma, « New Hampshire Opens Criminal Probe into AI Calls Impersonating Biden », *The Washington Post,* 6 février 2024, disponible sur : www.washingtonpost.com. La Commission fédérale des communications (FCC) américaine est également entrée dans la danse pour rendre illégales ces pratiques. Lire « AI-generated Voices in Robocalls Can Deceive Voters: The FCC Just Made Them Illegal », *Politico*, 8 février 2024, disponible sur www.politico.com.

délétère[60]. Par ailleurs, en dépit d'un volontarisme de façade[61], la désinvolture des entreprises du secteur en matière de gestion des contenus manipulés est également régulièrement pointée du doigt[62]. Dans ce contexte, la nécessité de labelliser les contenus créés par ou à l'aide d'IA générative devient de plus en plus prégnante, sans que cela n'offre de garantie absolue face aux techniques de contournement[63].

Car le recours à l'IA générative se fait progressivement plus complexe et plus subtil, permettant de capter une plus large audience et de contrer les efforts de détection mis en place par les plateformes[64], en recourant notamment à des vecteurs indirects (influenceurs ou leaders de communautés en ligne) qui republient sans le savoir des contenus générés par IA. De nouvelles pratiques en matière d'attaques informationnelles « complexes » sont ainsi d'ores et déjà en train d'émerger, telle la diffusion de faux journaux télévisés[65]. L'IA générative vient de fait renforcer le *sharp power* et appuyer les intérêts stratégiques des États qui souhaitent pénétrer les sphères informationnelles de leurs adversaires pour les influencer et y modifier les perceptions[66].

60. L'équipe de campagne de Donald Trump ne semble pas s'être alarmée outre mesure que des sympathisants de ce dernier n'hésitent pas à créer de toutes pièces des photos du candidat entouré d'Afro-américains pour peser sur le vote de ces derniers. Lire M. Spring, « Trump Supporters Target Black Voters with Faked AI Images », BBC, 4 mars 2024, disponible sur : www.bbc.com. Les démocrates et leurs soutiens n'hésitent pas non plus à recourir à l'IA, contribuant à en élargir les usages politiques. Lire M. Chaterjee et M. Fernandez, « An Arms Race Forever' as AI Outpaces Election Law », *Politico*, 7 février 2024, disponible sur www.politico.com ; R. Kern, M. Chaterjee et M. Fernandez, « A Democratic Campaign Deploys the First Synthetic AI Caller », *Politico*, 12 décembre 2023, disponible sur : www.politico.com.

61. Google, Meta, Microsoft, OpenAI, TikTok et Adobe ont encore récemment signé un accord pour renforcer la 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, disponible sur : www.washingtonpost.com. Cela n'empêche pas les développements en interne d'outils potentiellement dévastateurs, à l'image du projet VASA-1 de Microsoft Asia, qui permettrait de réaliser des *deepfakes* à partir d'une seule photo et un extrait audio. Lire B. Edwards, « Microsoft's VASA-1 Can Deepfake a Person with One Photo and One Audio Track », Ars Technica, 19 avril 2024, disponible sur : htpps://arstechnica.com.

62. C. Zakrzewski, « ChatGPT Breaks Its Own Rules on Political Messages », *The Washington Post,* 28 août 2024, disponible sur : www.washingtonpost.com ; N. Nix, « Oversight Board Rebukes Meta's Policies After Altered Biden Video Spreads », *The Washington Post,* 5 février 2024, disponible sur : www.washingtonpost.com.

63. N. Clegg, « Labeling AI-Generated Images on Facebook, Instagram and Threads », Meta, 6 février 2024, disponible sur : about.fb.com ; « Many AI Researchers Think Fakes Will Become Undetectable », *The Economist*, 17 janvier 2024, disponible sur : www.economist.com.

64. « Shadow Play », l'une des dernières campagnes d'influence chinoise en date, ayant ciblé les États-Unis et leurs alliés (dont la France), s'est déroulée sur YouTube et recourait notamment à l'IA générative pour créer de toutes pièces les voix d'influenceurs fictifs, déroulant des contenus orientés favorables à Pékin. Lire J. Keast, « Shadow Play, A pro-China Technology and Anti-US Influence Operation Thrives on YouTube », Australian Strategic Policy Institute, décembre 2023, disponible sur : www.aspi.org.

65. Un journal télévisé généré par IA a ainsi été diffusé par des hackers pro-iraniens aux Émirats arabes unis. Lire D. Milmo, « Iran-backed Hackers Interrupt UAE TV Streaming Services with Deepfake News », *The Guardian*, 8 février 2024, disponible sur : www.theguardian.com.

66. Le concept de *sharp power*, soumis par deux chercheurs de la National Endownment for Democracy, prend notamment appui sur les travaux de Joseph Nye concernant les deux grandes formes de la puissance que sont la coercition (*hard power*) et l'influence (*soft power*). Le *sharp power* désigne

Les tactiques employées pourraient recouvrir divers objectifs : faire pression sur les personnels encadrant les élections *via* des pratiques systématiques de divulgation malveillante d'informations personnelle (*doxing*) ; tromper les électeurs en utilisant de faux messages ou des images modifiées de candidats aux élections, en propageant des informations erronées sur les emplacements et les heures d'ouverture des bureaux de vote ; tenter de disqualifier un scrutin en diffusant de fausses images ou vidéos de fraudes, etc.[67]

Les tentatives sont déjà multiples, allant de la fabrication de contenus à l'usurpation d'identité en contexte électoral. Les États occidentaux sont ainsi régulièrement ciblés par la Russie, la Chine ou l'Iran, et les États-Unis constituent à ce titre l'un des laboratoires des pratiques de désinformation par IA[68]. Mais ils ne sont pas seuls : les dernières élections en Slovaquie, en Argentine, au Bangladesh ou encore en Indonésie ont elles aussi été marquées par des usages agressifs de ces technologies[69], et la scène politique indienne semble également s'engager sur cette pente[70]. À l'opposé du spectre, les États autoritaires tentent comme à leur habitude de verrouiller leur propre espace informationnel, n'ayant guère tardé à interdire les *chatbots* étrangers et à imposer la censure de rigueur à ceux produits par leurs entreprises nationales[71]. De fait, l'IA y sert plutôt d'outil opportun pour réduire un peu plus les espaces de liberté en ligne[72].

Entre renforcement de risques existants et nouveaux dangers

L'avènement de l'IA générative représente à la fois une opportunité et un risque pour la cybersécurité des systèmes d'information. Si les acteurs cyberoffensifs peuvent recourir aux LLM pour effectuer des tâches *a priori* non

notamment la capacité des États autoritaires à manœuvrer dans une zone grise entre *hard power* et *soft power,* en recourant à la manipulation et la subversion pour saper les fondements des systèmes démocratiques (processus électoraux, espaces d'information de débat publics, confiance dans les institutions, etc.). Lire C. Walker et J. Ludwig, « The Meaning of Sharp Power: How Authoritarian States Project Influence », *Foreign Affairs,* 16 novembre 2017, disponible sur www.foreignaffairs.com.

67. J. Easterly, S. Schwabb et C. Conley, « Artificial Intelligence's Threat to Democracy », *Foreign Affairs*, 3 janvier 2024, disponible sur : www.foreignaffairs.com.

68. *Ibid.*

69. O. Solon, « Trolls in Slovakian Election Tap AI Deepfakes to Spread Disinfo », *Bloomberg*, 29 septembre 2023, disponible sur : www.bloomberg.com ; J. Nicas et L. Cholakian Herrera, « Is Argentina the First A.I. Election? », *The New York Times*, 15 novembre 2023, disponible sur : www.nytimes.com ; D. Muller, « Deepfakes for $24 a Month: How AI Is Disrupting Bangladesh's Election », *Financial Times*, 14 décembre 2023, disponible sur : www.ft.com ; K. Lamb, F. Potkin et A. Teresia, « Generative AI May Change Elections This Year: Indonesia Shows How », *Reuters*, 9 février 2024, disponible sur : www.reuters.com.

70. N. Christopher, « 'Inflection Point': AI Meme Wars Hit India Election, Test Social Platforms », *Al Jazeera*, 8 mars 2024, disponible sur : www.aljazeera.com.

71. A. Funk et A. Shahbaz, « AI Chatbots Are Learning to Spout Authoritarian Propaganda », Wired, 4 octobre 2023, disponible sur : www.wired.com.

72. A. Funk, A. Shahbaz et K. Vesteinsson, « The Repressive Power of Artificial Intelligence », Freedom House, 2023, disponible sur : www.freedomhouse.org.

stratégiques (traduction automatique, correction de code informatique, digestion d'éléments techniques, etc.), ces modèles leur offrent de fait l'opportunité de renforcer leurs capacités en matière d'ingénierie sociale, d'usurpation et de compromission – en améliorant notamment le ciblage, la qualité et la quantité des tentatives d'hameçonnage[73] – mais aussi de générer du code plus rapidement et de manière plus « créative » – comme semblent s'y adonner des groupes chinois, iraniens, nord-coréens ou russes[74].

Selon le National Cyber Security Center britannique, les LLM seraient particulièrement vulnérables à deux types d'attaques : les instructions malveillantes (*prompt injections*), qui visent à manipuler des éléments du modèle *via* l'interface de commande, et les corruptions de jeux de données (*data poisoning*), qui peuvent advenir en amont et pendant la phase d'entraînement des modèles. Dans la mesure où les LLM sont entraînés à partir d'immenses jeux de données ouvertes et sont de plus en plus utilisés pour transmettre des données à des applications et services tiers, ces attaques par corruption des jeux de données constituent un risque non négligeable.

Il est également probable que l'IA générative facilitera tout autant l'insertion de cybercriminels novices qu'elle créera des fenêtres de tir pour les hackers opportunistes proposant leurs services au plus offrant, renforçant la prolifération déjà problématique d'armes cyber. Dans ce contexte, la menace mondiale en matière de rançongiciels devrait continuer de croître au cours des prochaines années, de même que les attaques par déni de service distribué (DDoS) – l'IA pouvant contribuer à optimiser la coordination et la synchronisation des attaques *via* des réseaux de machines infectées (*botnets*). Si le crime organisé profite déjà de ces nouveaux usages[75], les acteurs étatiques aux capacités cyber solidement établies devraient néanmoins rester les mieux placés pour tirer parti de cette évolution – dont certains doutent qu'elle bouleverse véritablement la donne existante[76].

Bien qu'aucune cyberattaque d'ampleur *via* le recours à l'IA générative ne soit publiquement connue jusqu'ici, les acteurs du secteur, à l'image de Microsoft et OpenAI, communiquent sur leur vigilance spécifique à l'égard des groupes de hackers étatiques et des pratiques qu'ils ont pu observer jusqu'à présent. États comme entreprises tentent de se préparer en amont à ces risques[77] et cherchent à renforcer la coordination de leurs moyens de

73. J. Hazell, « Spear Phishing with Large Language Models », Oxford Internet Institute, 14 décembre 2023, disponible sur : www.governance.ai.
74. « Staying Ahead of Threat Actors in the Age of AI », Microsoft, 14 février 2024, disponible sur : www.micrsoft.com.
75. D. Larousserie, « Comment les chatbots ont été gangrenés par l'industrie du crime organisé », *Le Monde*, 13 février 2024, disponible sur : www.lemonde.fr.
76. M. Untersinger, « Ni fin du monde ni panacée : l'intelligence artificielle n'a pas révolutionné la cybersécurité », *Le Monde*, 28 mars 2024, disponible sur : www.lemonde.fr.
77. L'administration américaine exige davantage de redevabilité des entreprises d'IA en matière de tests de sécurité, quand Google renforce par exemple ses programmes de recherche de vulnérabilités. Lire J. Boak, « AI Companies Will Need to Start Reporting Their Safety Tests to the US Government »,

cyberdéfense, à l'image de la création par la National Security Agency (NSA) américaine d'un Artificial Intelligence Security Center en septembre 2023[78].

Au-delà du seul domaine cyber, d'autres types de risques sont parfois mentionnés comme pouvant être renforcés par l'IA générative. À mesure que toujours plus de services fondés sur cette dernière verront le jour, certains usages poseront probablement des défis de nature sécuritaire pour nos sociétés fortement numérisées. Ainsi de l'usurpation d'identité, rendue potentiellement plus aisée par la génération réaliste de fausses photos de cartes d'identité[79].

Par ailleurs, des inquiétudes légitimes peuvent exister quant aux potentiels impacts de l'IA générative sur les menaces biologiques et chimiques[80]. Mais il faut garder à l'esprit que ce sujet est notamment poussé par les long-termistes, dans des perspectives de dissolution des efforts de régulation à court terme et de restriction d'accès au marché. De premiers travaux sur ces risques tendent à minorer les préoccupations, dans la mesure où le recours aux LLM ne permettrait pas – jusqu'ici – d'effectuer de saut qualitatif dans un domaine qui nécessite en tout état de cause l'accès à des laboratoires et des équipements avancés[81].

L'enjeu demeure ici de ne pas céder aux sirènes des « altruistes efficaces » qui ont fait de ce sujet leur cheval de bataille, quand bien même la communauté scientifique demeure divisée sur cette question, comme en témoignent certaines prises de position appelant à une étroite régulation de l'IA dans le champ biologique[82]. Ces recherches préliminaires[83] ont donc vocation à être poursuivies et approfondies pour apprécier plus finement ce risque, notamment à l'aune de futures applications en matière de conception de molécules et systèmes biologiques complexes tels que les génomes (*biodesign*).

Associated Press, 29 janvier 2024, disponible sur : https://apnews.com ; C. Page « Google Adds Generative AI Threats to Its Bug Bounty Program », *Techcrunch*, 26 octobre 2023, disponible sur https://techcrunch.com.

78. J. Clark, « AI Security Center to Open at National Security Agency », Département de la Défense américain, 28 septembre 2023, disponible sur : www.defense.gov.

79. J. Kox, « Inside the Underground Site Where "Neural Networks" Churn Out Fake Ids », *404 Media*, 5 février 2024, disponible sur : www.404media.co.

80. Cette préoccupation, mentionnée notamment dans le décret présidentiel sur l'IA signé par Joe Biden en octobre 2023, avait pris de l'ampleur à la suite des déclarations alarmistes de Dario Amodei (Anthropic) au Sénat américain.

81. Anthropic, OpenAI et la RAND Corporation ont notamment commencé à explorer cette question. Lire « Frontier Threats Red Teaming for AI Safety », *Anthropic*, 26 juillet 2023, disponible sur : www.anthropic.com ; « Building an Early Warning System for LLM-aided Biological Threat Creation », OpenAI, 31 janvier 2024, disponible sur : https://openai.com ; C. A. Mouton, C. Lucas et E. Guest, « The Operational Risks of AI in Large-Scale Biological Attacks », RAND, 25 janvier 2024, disponible sur : www.rand.org.

82. C. Metz, « Dozens of Top Scientists Sign Effort to Prevent A.I. Bioweapons », *The New York Times*, 8 mars 2024, disponible sur : www.nytimes.com.

83. S. Carter, N. Wheeler, S. Chwalek, C. Isaac et J. Yassif, « The Convergence of Artificial Intelligence and the Life Sciences », Nuclear Threat Initiative, 30 octobre 2023, disponible sur : www.nti.org.

Open source vs. concentration du marché

Les modèles de fondation (*foundation models* ou *General-purpose AI models*), qui permettent d'accomplir des tâches variées et constituent l'épine dorsale à partir de laquelle des modèles secondaires et des applications sont ensuite déployés, sont aujourd'hui développés par une poignée d'entreprises seulement. En raison d'un coût d'entrée très élevé, le marché récompense les primo-entrants et présente dès lors une tendance à la concentration[84]. Celle-ci se voit renforcée par les logiques de puissance financière à l'œuvre : si les sociétés de capital-risque semblent avoir momentanément marqué le pas[85], les investissements massifs des *Big Tech* dans les entreprises du secteur – à l'image de Google avec Deepmind, Microsoft avec OpenAI ou encore Amazon avec Anthropic[86] – ont contribué à une forte concentration du marché à leur profit.

Les ressources stratégiques (données, semi-conducteurs, puissance de calcul, talents…) se trouvent aujourd'hui regroupées entre les mains de ces acteurs, qui sont de fait en capacité de peser de tout leur poids sur les grandes orientations du secteur[87]. Les moyens dont ils disposent leur permettent ainsi de provoquer une véritable fuite des cerveaux qui déleste les universités, centres de recherche et autres institutions publiques des meilleurs éléments[88], tandis que leurs politiques de sécurité contribuent à entraver la recherche académique[89]. Cette situation suscite des inquiétudes légitimes dans la mesure où, comme le souligne Benoît Cœuré, président de l'Autorité de la concurrence française, « l'IA est la première innovation de rupture se produisant dans un paysage où les entreprises déjà les plus puissantes contrôlent les capacités pour la développer[90] ». Ce d'autant plus que ces géants ont un casier judiciaire chargé en matière de pratiques anticoncurrentielles et monopolistiques, passif qui a probablement incité au

84. J. Vipra et A. Korinek, « Market Concentration Implications of Foundation Models: The Invisible Hand of ChatGPT », Brookings, 7 septembre 2023, disponible sur : www.brookings.edu.
85. G. Nahon, « L'intelligence artificielle générative risque de plus en plus d'échapper au capital-risque, au profit des Big Tech », *Le Monde,* 5 février 2024, disponible sur : www.lemonde.fr.
86. E. Ludlow, M. Day et D. Bass, « Amazon to Invest Up to $4 Billion in AI Startup Anthropic », *Bloomberg*, 25 septembre 2023, disponible sur : www.bloomberg.com.
87. A. Kak, S. Myers West et M. Wittaker, « Make No Mistake, AI Is Owned by Big Tech », *MIT Technology Review,* 5 décembre 2023, disponible sur : www.technologyreview.com. Ceci crée des tensions au sein des entreprises du secteur de taille plus modeste – à l'image de StabilityAI – qui souffrent à la fois du débauchage et de divergences internes sur l'orientation à suivre face à la concentration du marché. Lire T. Warren, « Stability AI CEO Resigns to "Pursue Decentralized AI" », *The Verge*, 23 mars 2024, disponible sur : www.theverge.com.
88. N. Nix, C. Zakrzewski, G. De Vynck, « Silicon Valley Is Pricing Academics Out of AI Research », *The Washington Post*, 10 mars 2024, disponible sur : www.washingtonpost.com. En 2020, 70 % des doctorants américains en IA étaient recrutés par le secteur privé contre 20 % en 2004. Lire N. Ahmed et N. Thomson, « What Should Be Done about the Growing Influence of Industry in AI Research? », Brookings, 5 décembre 2023, disponible sur : www.brookings.edu.
89. N. Tiku, « Top AI Researchers Say OpenAI, Meta and More Hinder Independent Evaluations », *The Washington Post*, 5 mars 2024, disponible sur : www.washingtonpost.com.
90. A. Piquard, « Intelligence artificielle : mobilisation contre la domination annoncée des géants du numérique », *Le Monde*, 17 janvier 2024, disponible sur : www.lemonde.fr.

lancement rapide d'enquêtes préliminaires[91]. Dès lors, les appels à une politique non seulement de régulation mais aussi d'investissements publics dans l'IA pour équilibrer le poids pris par les acteurs privés se multiplient[92].

Dans ce contexte, l'*open source* représente à la fois l'espoir d'une alternative nécessaire et complémentaire aux modèles dominants[93], mais également une source d'inquiétude, notamment parce qu'il met des moyens d'ampleur stratégique à une large disposition, ce qui induit des enjeux de sécurité. Contrairement aux entreprises propriétaires, les développeurs de modèles « ouverts » ont de fait une capacité limitée à restreindre l'utilisation par des acteurs malveillants de leurs modèles une fois ceux-ci diffusés. Il faut toutefois souligner qu'ils échappent dans le même temps plus facilement aux injonctions d'une course commerciale poussant les acteurs propriétaires à préserver leur avantage concurrentiel en sortant rapidement des modèles pas toujours pleinement aboutis ni sécurisés.

La concentration peut également aboutir à l'apparition de défaillances d'ordre systémique, dans la mesure où les modèles propriétaires dominants sont aujourd'hui les plus usités. La question s'inscrit donc pour une large partie dans un débat plus large et préexistant sur la (cyber)sécurité de l'*open source*, mais entraîne ici des crispations plus fortes du fait du potentiel encore mal cerné des technologies d'IA générative[94]. Alors que l'IA contribue d'ores et déjà à une meilleure sécurisation de l'*open source*[95], de premiers travaux tendent à démontrer que ce qui vaut pour le logiciel libre vaut pour les modèles de fondation « ouverts » : le risque sécuritaire serait d'ordre limité, pour peu qu'il fasse l'objet d'un encadrement approprié et proportionné – ne pesant pas uniquement sur les développeurs[96]. Ces recherches appellent là encore à être approfondies pour déterminer qui

91. Les autorités de la concurrence britanniques ont ainsi ouvert une enquête sur le rachat d'OpenAI par Microsoft et pourraient être rapidement imitées par la Commission européenne, tandis que l'Afrique du Sud a annoncé enquêter sur les modèles de Google et Meta, qui limiteraient la capacité des acteurs journalistiques locaux à dégager des revenus. Lire D. Milmo, « UK Watchdog to Examine Microsoft's Partnership with OpenAI », *The Guardian*, 8 décembre 2023, disponible sur *:* www.theguardian.com ; « Commission Launches Calls for Contributions on Competition in Virtual Worlds and Generative AI », Communiqué de presse, Commission européenne, 9 janvier 2024, disponible sur : https://ec.europa.eu ; L. Prinsloo, « Big Tech, AI Models to Face Antitrust Inquiry in South Africa », *Bloomberg*, 17 octobre 2023, disponible sur : www.bloomberg.com.

92. Comme le stipule Marietje Schaake, directrice de la politique internationale au Cyber Policy Center de Standford, « l'IA est trop importante pour faire l'objet d'un monopole ». Lire M. Schaake, « AI Is Too Important to Be Monopolised », *Financial Times*, 12 décembre 2024, disponible sur : www.ft.com.

93. Avec les mêmes arguments que pour le code ouvert et le logiciel libre : transparence et accessibilité accrue, redevabilité collective, coût proportionnel aux usages, évaluation et sécurisation par les pairs ou les tiers, stimulation de l'innovation, etc.

94. T. Shevlane et A. Dafoe, « The Offense-Defense Balance of Scientific Knowledge: Does Publishing AI Research Reduce Misuse? », Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES'20), 7-8 février 2020, disponible sur : https://arxiv.org.

95. Notamment en aidant à détecter des vulnérabilités dans les multitudes de projets *open source* plus ou moins bien maintenus. Lire D. Liu, J. Metzman et O. Chang, Google OS Security Team, « AI-Powered Fuzzing: Breaking the Bug Hunting Barrier », *Google Security Blog*, 16 août 2023, disponible sur : https://security.googleblog.com.

96. R. Bommasani, S. Kapoor, K. Klyman *et al.*, « Considerations for Governing Open Foundation Models », HAI Policy & Society, décembre 2023, disponible sur : https://hai.stanford.edu.

des modèles propriétaires ou « ouverts » parvient le mieux à réduire les risques auxquels ils sont collectivement exposés.

Partiellement encouragé jusqu'ici par les autorités américaines et européennes, le déploiement de modèles ouverts souffre de l'attitude ambivalente des grandes entreprises du secteur à l'égard de l'*open source*[97]. D'un côté, des développeurs comme Meta, BigScience, Stability AI, Hugging Face, Mistral, Aleph Alpha, Together AI ou EleutherAI publient des modèles ouverts à des degrés divers (avec ou sans accès aux données d'entraînement, avec ou sans restriction d'utilisation), souvent assorti d'« *open washing*[98] ». De l'autre, des modèles fermés (Flamingo de Deepmind, Gemini et PaLM2 de Google, Claude d'Anthropic) ou semi-fermés (ChatGPT d'OpenAI) comptent aussi parmi les plus réputés et usités. Les deux types d'orientation peuvent coexister comme chez Mistral, qui après avoir initialement produit des modèles ouverts, propose aussi à présent une offre « fermée » (« Mistral Large ») résultant de son partenariat avec Microsoft. En fonction de la stratégie commerciale qui lui est propre, chacun tente de jouer sa partition. Meta, qui pourrait utilement profiter de son bon positionnement dans l'IA pour renforcer l'efficacité et l'attractivité de ses plateformes (Facebook, Instagram, Whatsapp, etc.), se targue opportunément d'être le champion de l'*open source*, s'attirant les faveurs des talents attachés à la préservation d'une forme d'ouverture. Anthropic ou OpenAI, qui cherchent aujourd'hui à tirer profit des modèles généralistes qu'ils développent face à la plus grande spécialisation sectorielle de modèles *open source*, insistent quant à eux sur les potentiels risques de ces derniers en matière de sécurité[99].

Ces récits et visions concurrentes sont ensuite activement relayés auprès des régulateurs par des groupes de pression (AI Alliance vs. Frontier Model Forum), dans l'espoir de peser sur les orientations futures[100]. L'enjeu pour ces acteurs est à la fois de protéger et rentabiliser leurs innovations propriétaires qui sont régulièrement imitées voire rattrapées les années suivantes par les progrès des communautés *open source*, mais aussi de maintenir ces dernières dans leur giron pour mieux s'en nourrir. De l'autre côté du spectre, les acteurs plus « modestes » sont davantage enclins à

97. A. Pannier, « Sources d'influence. Enjeux économiques et géopolitiques des logiciels *open source* », *Études de l'Ifri*, Ifri, décembre 2022.
98. E. Gent, « The Tech Industry Can't Agree on What Open-source AI Means. That's a Problem », *MIT Technology Review*, 25 mars 2024, disponible sur : www.technologyreview.com.
99. Voir notamment les déclarations respectives de Mark Zukerberg (Meta) et Dario Amodei (Anthropic). Lire J. Koetsier, « Meta to Build Open-Source Artificial General Intelligence for All, Zuckerberg Says », *Forbes*, 18 janvier 2024, disponible sur : www.forbes.com ; « Anthropic CEO: It's Hard to Keep Open Source AI Models Safe », YouTube, 22 novembre 2023, disponible sur : www.youtube.com.
100. L'AI Alliance, à l'instigation de Meta et IBM, regroupe des acteurs aussi divers qu'Oracle, AMD, Uber, Sony, Hugging Face, Stability AI, la Linux Foundation et divers partenaires universitaires en faveur de l'innovation et la science ouvertes dans le domaine de l'IA. De son côté, le Frontier Model Forum, qui réunit Anthropic, Google, Miscrosoft et OpenAI, souhaite soutenir les efforts pour la sécurisation et le développement « d'applications d'IA pour répondre aux besoins les plus pressants de la société ». Lire M. O'Brien, « AI's Future Could Be "Open-source" or Closed: Tech Giants Are Divided as They Lobby Regulators », Associated Press, 5 décembre 2023, disponible sur : https://apnews.com.

promouvoir l'ouverture car celle-ci leur permet de réduire leurs coûts d'entrée. Ils tentent eux aussi de faire entendre leur voix, à l'image des démarches entreprises par Hugging Face et d'EleutherAI auprès des autorités européennes[101].

Alors que l'IA *open source* ne pourra probablement pas à elle seule permettre une véritable déconcentration du marché, deux risques sérieux pèseront sur son développement à moyen terme, en plus des tentatives d'ores et déjà à l'œuvre pour la discréditer[102]. Le premier est celui de sa recaptation ultérieure (*enclosure*) par les grands acteurs privés, qui possèdent de fait un levier de taille car ils fournissent les infrastructures sur lesquelles tournent la majorité des modèles *open source,* et disposent des capitaux nécessaires à d'éventuels rachats.

Il faut y ajouter le risque d'un revirement des autorités américaines[103] : dans un contexte où une initiative commune avec la Chine pour assurer la gouvernance de l'IA *open source* a peu de chances de voir le jour, la volonté à Washington de maintenir un avantage compétitif pourrait *in fine* conduire à des velléités de limiter la production et l'accès à des modèles ouverts, sous couvert de sécurité nationale. Une alliance objective entre faucons de la politique américaine vis-à-vis de la Chine (*China hawks*) et entreprises propriétaires opposées à l'*open source* pourrait dès lors voir le jour, dans le sillage de lobbys qui lient déjà ces deux questions[104].

101. A. Coll, « Supporting Open Source and Open Science in the EU AI Act », 25 juillet 2023, disponible sur : https://huggingface.co.

102. Certains, à l'image du Futur of Life Institute, n'hésitent pas à brandir le risque nucléaire pour obtenir un strict encadrement de l'IA open source. Lire Future of Life Institute, « Artificial Escalation », YouTube, 17 juillet 2023, disponible sur : www.youtube.com. Cet organisme créé en 2014 par un universitaire du MIT, relativement inconnu avant de publier la lettre ouverte pour un moratoire sur l'IA, est notamment soutenu par E. Musk, V. Buterin (Ethereum) et J. Tallinn (Skype). Il est visiblement aligné sur les positions – si ce n'est les intérêts – des *effective altruists*, qu'il défend en particulier à Bruxelles, et mène des actions de lobbying sur des problématiques long-termistes (IA générale, risques existentiels et futur de l'humanité). Lire G. Volpicelli, « Stop the Killer Robots! Musk-backed Lobbyists Fight to Save Europe from Bad AI », *Politico*, 24 novembre 2022, disponible sur : www.politico.com.

103. Il faut souligner que jusqu'à présent le gouvernement américain semble vouloir soutenir l'écosystème de l'IA *open source*, que ce soit en en mentionnant l'importance dans ses décrets présidentiels, ou en créant un National AI Research Resource (NAIRR) pilot, afin de fournir des capacités de calcul et d'accès aux données aux milieux académiques. Mais l'élection présidentielle pourrait bien rebattre les cartes dans les mois à venir.

104. K. Xiu, « Washington Is Using China to Destroy Open Source », *Interconnected*, 13 octobre 2023, disponible sur : https://interconnected.blog. Pour autant, les efforts des lobbys – à commencer par l'AI Alliance – qui soutiennent l'*open source* s'appuient eux aussi sur le narratif du dépassement technologique chinois pour convaincre Washington de leur laisser davantage de champ et semblent en bonne voie d'y parvenir. Lire B. Bordelon, « In DC, a New Wave of AI Lobbyists Gains the Upper Hand », *Politico*, 12 mai 2024, disponible sur : www.politico.com.

L'environnement, le grand perdant

À l'heure où la COP28 souhaiterait « mettre l'IA au service de l'action climatique[105] », l'impact environnemental du développement de l'IA générative, souvent minoré ou passé sous silence, est en train de devenir un enjeu prégnant. En effet, le recours aux modèles induit une importante consommation de ressources, à la fois en amont et en aval de leur entraînement : fabrication de puces, stockage de données, entraînement des modèles, requêtes des utilisateurs et données ainsi générées ont des conséquences sur les plans physique, hydrique, énergétique et par conséquent climatique. Le poids écologique de ces technologies pourrait dès lors représenter un problème majeur[106].

Les estimations varient en fonction des études, mais la consommation de l'industrie semble appelée à croître démesurément[107], à tel point que certains leaders du secteur tirent d'ores et déjà la sonnette d'alarme[108]. La production de modèles toujours plus gourmands en données et donc en énergie, en dépit de tentatives d'en améliorer l'efficience, constitue de fait une fuite en avant préoccupante, alors même que les externalités positives semblent parfois très limitées[109] et que les systèmes énergétiques sont déjà mis à rude épreuve, y compris dans les pays les plus développés[110]. Ce besoin en énergie fait

105. « Mettre l'intelligence artificielle au service de l'action climatique dans les pays en développement, voici le défi lancé à la COP28 », United Nations Climate Change, 9 décembre 2023, disponible sur : https://unfccc.int.

106. L'impact carbone exponentiel du secteur est d'ores et déjà préoccupant. Lire A. de Vries, « The Growing Energy Footprint of Artificial Intelligence », *Sciences Direct*, vol. 7, n° 10, 2023, p. 2191-2194, disponible sur : www.sciencesdirect.com. Les IA génératives, par définition généralistes, seraient trente fois plus gourmandes en énergie que les IA spécialisées (dédiées à une seule tâche). La génération d'une image consommerait ainsi autant que la charge complète d'un smartphone, et la formulation d'un prompt équivaudrait à consommer 50 centilitres d'eau, en sachant que ces calculs n'ont pas été réalisés sur les modèles les plus énergivores. Lire A. S. Luccioni, Y. Jernite et E. Strubell, « Power Hungry Processing: Watts Driving the Cost of AI Deployment? », *Arxiv*, 28 novembre 2023, disponible sur : https://arxiv.org.

107. E. Kolbert, « The Obscene Energy Demands of A.I. », *The New Yorker*, 9 mars 2024, disponible sur : www.newyorker.com.

108. À commencer par Sam Altman (OpenAI), qui estime que le futur de son industrie, hautement consommatrice en électricité, sera étroitement corrélé à la possibilité d'avancées majeures en matière de production énergétique (fusion nucléaire et stockage des ressources produites par les énergies renouvelables). Lire V. Tangermann, « Sam Altman Says AI Using Too Much Energy, Will Require Breakthrough Energy Source », *Futurism*, 17 janvier 2024, disponible sur : https://futurism.com. En retour, d'autres voix, à l'image de Sasha Luccioni (Hugging Face), appellent à réduire dès à présent l'impact de l'IA générative et ses usages immodérés.

109. On peut par exemple légitimement s'interroger sur la portée réelle de Genie, modèle de DeepMind (Google) qui a jusqu'ici pour fonction de produire des mini-jeux vidéo à partir de simples croquis. Lire W. D. Heaven, « Google DeepMind's New Generative Model Makes Super Mario–like Games from Scratch », *MIT Technology Review*, 29 février 2024, disponible sur : www.technologyreview.com.

110. C'est notamment le cas aux États-Unis, où dans le sillage du développement de l'IA le réseau peine à faire face à l'accroissement de la demande, laquelle menace les ambitions de réduction d'émissions. Lire E. Halper, « Amid Explosive Demand, America Is Running Out of Power », *The Washington Post*, 7 mars 2024, disponible sur : www.washingtonpost.com ; C. Mims, « AI Is Ravenous for Energy: Can It Be Satisfied? », *The Wall Steet Journal*, 15 décembre 2023, disponible sur : www.wsj.com ; A. Freedman, « AI Advances May Frustrate U.S. Climate Goals as Electric Demand Surges », *Axios*, 26 avril 2024, disponible sur : www.axios.com.

notamment envisager des « solutions » à haut risque, comme la construction de réacteurs nucléaires au sein de centres de données spécialisés[111].

Alors que près de 20 % de la capacité totale des centres de données est consacrée à répondre aux besoins de l'IA, la création de centres dédiés (combinant gestion des données nécessaires à l'entraînement des modèles mais aussi à l'hébergement des données produites par l'IA[112]) pourrait faire peser un risque environnemental accru sur les territoires locaux, notamment lorsque ceux-ci sont déjà soumis à un stress hydrique et/ou énergétique important[113]. Les débats à ce sujet devraient donc s'intensifier à l'avenir, d'autant plus que les promoteurs de bon nombre de projets d'implantation semblent avancer masqués[114]. Des divergences profondes pourraient apparaître quant à la localisation durable ou non de tels centres, ou quant au juste partage des ressources entre une industrie cherchant à sécuriser et négocier ses approvisionnements en amont d'un côté, et le reste des activités et besoins propres à chaque territoire de l'autre[115]. Ceci ne manquerait pas d'élargir le fossé grandissant entre techno-enthousiastes et techno-sceptiques, bien visible dans le déploiement de la 5G dans les sociétés occidentales. La France ne sera probablement pas épargnée, dans la mesure où les incitations à bâtir de nouveaux centres de données devraient se multiplier, alors que de premières conséquences à l'échelle locale se font sentir[116].

111. On ne peut que s'inquiéter de cette volonté d'élargir la disponibilité et la manipulation des technologies nucléaires, qui requiert un niveau d'expertise et de sûreté extrêmement exigeant, dans un secteur pour le moment sous-régulé. Lire M. Dempsey, « Future Data Centres May Have Built-in Nuclear Reactors », BBC, 15 février 2024, disponible sur : www.bbc.com.

112. À l'image du centre de données annoncé par Meta dans l'Indiana. Lire K. Wagner, « Meta Is Building New $800 Million AI-Focused Data Center in Indiana », *Bloomberg*, 25 janvier 2024, disponible sur : www.bloomberg.com.

113. C'est notamment le cas avec un data center de Microsoft en Arizona, de Google en Uruguay et de Meta à Talavera de la Reina (Espagne). Lire K. Hao, « AI Is Taking Water from the Desert », *The Atlantic*, 1er mars 2024, disponible sur : www.theatlantic.com ; G. Livingstone, « "It's Pillage": Thirsty Uruguayans Decry Google's Plan to Exploit Water Supply », *The Guardian,* 11 juillet 2023, disponible sur : www.theguardian.com ; G. de Pierrefeu, « L'intelligence artificielle est-elle l'alliée incontournable, ou l'ennemie incontestable, de nos préoccupations environnementales ? », *Le Monde*, 16 janvier 2024, disponible sur : www.lemonde.fr.

114. Une trentaine de centres de données auraient ainsi été discrètement implantés par les géants technologiques « sous fausse identité ». Lire D. Jeans, « Data In The Dark: How Big Tech Secretly Secured $800 Million in Tax Breaks for Data Centers », *Forbes*, 19 août 2021, disponible sur : www.forbes.com.

115. C. Criddle et K. Bryan, « AI Boom Sparks Concern over Big Tech's Water Consumption », *Financial Times*, 24 février 2024, disponible sur : www.ft.com ; P. Sisson, « A.I. Frenzy Complicates Efforts to Keep Power-Hungry Data Sites Green », *The New York Times*, 29 février 2024, disponible sur : www.nytimes.com.

116. O. Pinaud, « La Commission de l'intelligence artificielle veut faciliter l'installation de centres de données en France », *Le Monde*, 13 mars 2024, disponible sur : www.lemonde.fr ; A. Piquard, « L'explosion de la demande d'électricité liée à l'IA a déjà des conséquences locales », *Le Monde*, 8 février 2024, disponible sur : www.lemonde.fr.

Vers un renforcement des inégalités et du « clivage Nord-Sud »

Les débats sur les impacts sociétaux de l'IA générative – notamment en matière de destruction d'emplois – sont déjà vifs dans les pays occidentaux. L'essor de ces technologies fait craindre une augmentation rapide d'inégalités déjà criantes, qui pourrait s'accompagner de risques sociaux-politiques non négligeables, pour peu que les sociétés s'avèrent incapables de proposer des solutions durables aux franges « déclassées » de leur population[117]. Contrairement à la *doxa* estimant que les pays émergents auront l'opportunité d'effectuer un saut technologique et bénéficieront d'autant plus des apports de l'IA que leurs sociétés reposent moins sur une économie du savoir fortement automatisée, il semble que ceux-ci puissent être à terme les plus exposés aux bouleversements sociaux à venir. La concurrence à l'emploi et la pression exercée par l'IA sur des métiers à faible valeur ajoutée y sont de fait plus fortes, en raison de la moindre protection sociale, de l'absence de recours possibles et de la plus grande difficulté à accéder à ces technologies – et donc à en tirer parti[118].

Qui plus est, ce sont ces pays qui comptent le plus de travailleurs du clic nécessaires à l'entraînement des modèles[119], dont les conditions de vie précaires ne peuvent manquer d'alimenter un ressentiment au regard des richesses drainées par l'industrie[120]. L'asymétrie de la captation de données, de même que l'iniquité de la répartition des coûts et bénéfices entre une IA alimentée pour partie physiquement – en ressources minières – et humainement par les travailleurs du Sud[121], développée par et pour les

117. Une étude du Fonds monétaire international estime ainsi que 60 % des emplois dans les pays les plus développés pourraient être transformés, affectés voire remplacés par l'IA, contre 20 à 40 % dans les pays émergents. Mais elle souligne également que ces derniers auront moins d'opportunités de bénéficier des apports de l'IA, faute d'infrastructures et de main-d'œuvre qualifiée, ce qui pourrait conduire à creuser les inégalités. Lire K. Georgieva, « AI Will Transform the Global Economy: Let's Make Sure It Benefits Humanity », *IMF Blog,* 14 janvier 2024, disponible sur : www.imf.org. Il faut également rappeler que l'adoption de l'IA sert parfois de prétexte pour opérer des coupes budgétaires et des licenciements massifs, notamment dans le secteur technologique. Lire M. Morrone, « "AI Made Us Do It" Is Big Tech's New Layoff Rationale », *Axios*, 18 janvier 2024, disponible sur : www.axios.com.

118. À titre d'exemple, les web-développeurs et designers sud-africains souffrent directement de cette mise en concurrence. Lire K. Mutandiro, « Free AI Tools Are Killing South Africa's Web Designer Job Market », *Rest of World*, 31 août 2023, disponible sur : https://restofworld.org.

119. Il faut rappeler que ces derniers sont par ailleurs en première ligne face à la violence des contenus dont ils ont la charge. Lire K. Hao et D. Seetharaman, « Cleaning Up ChatGPT Takes Heavy Toll on Human Workers », *Wall Street Journal*, 24 juillet 2023, disponible sur : www.wsj.com.

120. B. Perrigo, « OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic », *Time*, 18 janvier 2023, disponible sur : www.time.com.

121. Comme le stipule D. Björkegren, « on a beaucoup plus investi dans des applications destinées à mettre en relation des consommateurs riches avec des chauffeurs, des maisons de vacances et des repas préparés que dans des applications destinées à lier des agriculteurs de subsistance avec des marchés ou des enfants isolés à renouer avec l'apprentissage. L'innovation du secteur privé en matière d'IA est susceptible de transformer de nombreux secteurs, de l'éducation à la santé en passant par le droit. Mais pour exploiter tout le potentiel de cette technologie dans les pays en développement, il faudra proposer une vision élargie de ce qui est possible – et prêter une attention particulière aux personnes dont la vie pourrait en être

populations du Nord, créent un sentiment d'injustice qui contribue à renforcer les clivages préexistants. Ceci d'autant plus que la plupart des systèmes d'IA souffrent d'un profond défaut de représentativité – ayant été entraînés sur des données issues du monde développé, recueillies majoritairement auprès d'hommes ayant des revenus élevés, et généralement écrites en anglais – et véhiculent dès lors des stéréotypes susceptibles d'accroître les inégalités internes à chaque société mais aussi les fossés politico-culturels entre États[122]. L'Organisation mondiale de la santé (OMS) s'est ainsi fendue d'un avertissement quant aux potentiels impacts des technologies de santé dopées à l'IA dans les pays émergents, s'inquiétant notamment du manque de diversité des données d'entraînement pour servir efficacement les populations peu représentées, et du risque de mainmise sur le développement des LMM par le secteur privé, aux dépens de la recherche universitaire et des agences gouvernementales.

Face aux risques d'accroissement de ces multiples clivages sociopolitiques, les appels à la mise en place de garde-fous se multiplient, y compris au plus haut niveau[123]. Diverses propositions, tel l'établissement d'une taxe spécifique pour amortir les dégâts sociaux[124], semblent émerger progressivement. Jugeant le rapport risques-bénéfices de l'IA profondément asymétrique, d'autres voix réclament une gouvernance « technoprudentielle[125] », voire un « endiguement » social de l'IA[126], impliquant décideurs, industriels et société civile pour que nos sociétés puissent maintenir le contrôle sur ces technologies et en endiguer les externalités négatives.

bouleversée », *in* D. Björkegren, « Artificial Intelligence for the Poor: How to Harness the Power of AI in the Developing World », *Foreign Affairs,* 9 août 2023, disponible sur : www.foreignaffairs.com.

122. V. Türk, « How AI Reduces the World to Stereotypes », *Rest of World*, 10 octobre 2023, disponible sur : www.restofworld.org.

123. À l'image du Secrétaire général des Nations unies à Davos, pour qui « 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. Il nous faut également déployer des efforts systématiques pour améliorer l'accès à l'IA afin que les économies en développement puissent bénéficier de son énorme potentiel. Nous devons combler le fossé numérique au lieu de l'aggraver ». Lire L. Eliott, « Big Tech Firms Recklessly Pursuing Profits from AI, Says UN Head », *The Guardian*, 17 janvier 2024, disponible sur : www.theguardian.com.

124. Une idée notamment portée par l'ancienne élue européenne aujourd'hui chercheuse à Standford Marietje Schaake : M. Schaake, « It's Already Time to Think About an AI Tax », *Financial Times,* 8 janvier 2024, disponible sur : www.ft.com.

125. 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, disponible sur : www.foreignaffairs.com.

126. M. Suleyman, « Containment for AI: How to Adapt a Cold War Strategy to a New Threat », *Foreign Affairs*, 23 janvier 2024, disponible sur : www.foreignaffairs.com.

Une compétition protéiforme entre puissances et pour la puissance

Une course avant tout financière

Alors que le secteur technologique est en proie à un ralentissement économique, l'IA ne semble pas connaître la crise, son financement mondial ayant atteint près de 18 milliards de dollars d'investissements au troisième trimestre 2023[127]. Une *start-up* comme Anthropic (créée en 2021) aurait ainsi levé plus de 7 milliards de dollars l'an passé[128], avant même la sortie de son dernier modèle « Claude 3 », censé concurrencer ChatGPT4 et Gemini. Les États ne sont pas en reste et alimentent directement cette frénésie financière, annonçant régulièrement des efforts budgétaires colossaux, à l'image de l'Arabie saoudite, de la Corée du Sud ou du Canada ces derniers mois[129]. Les États-Unis et la Chine restent toutefois les plus gros pourvoyeurs de fonds publics, à hauteur de plusieurs milliards de dollars par an, et Washington semble même accélérer le rythme depuis deux ans[130].

Par ailleurs, la concurrence à laquelle se livrent les géants technologiques américains trouve de fait en l'IA générative un énième moteur, stimulant les investissements tous azimuts[131] : en 2022, Microsoft,

127. S. Mc Bride, « AI Funding Soars to $17.9 Billion While Rest of Tech Slumps », *Bloomberg*, 17 octobre 2023, disponible sur : www.bloomberg.com.

128. H. Field, « Anthropic, backed by Amazon and Google, Debuts Its Most Powerful Chatbot Yet », CNBC, 4 mars 2024, disponible sur : www.cnbc.com. Ces montants sont d'autant plus questionnables que la plupart de ces entreprises sont loin d'être rentables et enregistrent parfois des pertes records, à l'image d'OpenAI qui perdait 540 millions de dollars en 2022. Lire E. Woo et A. Efrati, « OpenAI's Losses Doubled to $540 Million as It Developed ChatGPT », *The Information*, 4 mai 2023, disponible sur : www.theinformation.com.

129. Ces États ont annoncé des efforts respectifs 40, 7 et 2 milliards de dollars sur les prochaines années, quand Abou Dabi a créé MGX, une société d'investissement dont l'objectif est de gérer 100 milliards de dollars d'actifs, dont une bonne part dans l'IA. Lire M. Farrell et R. Copeland, « Saudi Arabia Plans $40 Billion Push into Artificial Intelligence », *The New York Times*, 19 mars 2024, disponible sur : www.nytimes.com ; « South Korea to Invest $7 Billion in AI in Bid to Retain Edge in Chips », Reuters, 9 avril 2024, disponible sur : www.reuters.com ; « Canada's Trudeau Announces Package of AI Investment Measures », Reuters, 7 avril 2024, disponible sur : www.reuters.com ; B. Bartenstein, « Abu Dhabi Targets $100 Billion AUM for AI Investment Firm », *Bloomberg*, 11 mars 2024, disponible sur : www.bloomberg.com.

130. Notamment pour répondre aux besoins internes à l'administration. Lire Larson, J. Denford, G. Dawson et K. Desouza, « The Evolution of Artificial Intelligence (AI) Spending by the U.S. Government », Brookings, 26 mars 2024, disponible sur : www.brookings.edu.

131. Les géants technologiques investissent aujourd'hui davantage que les sociétés de capital-risque. Dans ce qui ressemble à une peur de manquer le tournant, Amazon et Google comptent injecter respectivement 4 et 2 milliards de dollars dans Anthropic, après que Microsoft a investi pas moins de 13 milliards dans OpenAI,

Amazon et Google ont ainsi abondé deux tiers des 27 milliards de dollars levés par les *start-ups* du secteur.

Au-delà de la captation d'une innovation encore incertaine et protéiforme, de tels investissements constituent à la fois une démonstration de puissance et une garantie réputationnelle, dans un milieu où le récit du gagnant raflant la mise (*winner takes all*) prédomine et peuple les imaginaires. Ils mettent ainsi d'autant plus sous pression ceux qui, à l'image d'Apple, sont vite perçus comme ayant pris du retard[132]. Cette compétition conduit même certaines entreprises à ordonner à leurs employés de ne pas recourir à des IA génératives tierces pour des raisons de « sécurité[133] » ou à restreindre l'accès de leurs modèles pour leurs concurrents[134], tandis que de premières poursuites judiciaires s'engagent.

Ces géants se sont également engagés dans des politiques de partenariats pour proposer des offres d'intégration de ces technologies plus ou moins structurées, avec pour objectif d'accroître de fait le marché en en facilitant l'accès[135]. Ils déploient dans le même temps de vastes et coûteux programmes de formation pour sensibiliser un plus large public à ces technologies[136]. Dans un cercle de retour sur investissement continu, les financements consentis sont en grande partie compensés par les revenus issus des infrastructures (en particulier de *cloud computing*) qu'ils allouent aux entreprises en IA. Qui plus est, ces investissements ont aussi pour conséquence de renforcer la confiance des investisseurs en leur leadership et, partant, leur capitalisation boursière. Ces manœuvres n'ont pas manqué d'attirer l'attention de la Federal Trade Commission (FTC) américaine, qui a déclenché une enquête sur ces pratiques[137].

racheté Inflection AI pour 1,3 milliard, et plus récemment placé 95 millions dans Figure et 15 millions dans MistralAI. Hugging Face est également soutenue par Microsoft et Amazon, et Humane par Microsoft et OpenAI. Lire H. Field et K. Leswing, « Generative AI "FOMO" Is Driving Tech Heavyweights to Invest Billions of Dollars in Startups », *CNBC*, 30 mars 2024, disponible sur : www.cnbc.com.

132. A. Tiley, « Apple Is Behind in AI—and Investors Are Getting Impatient », *Wall Street Journal*, 29 février 2024, disponible sur : www.wsj.com. Ceci a notamment conduit la firme à signer un contrat avec Google pour pouvoir équiper les iPhones du modèle Gemini.

133. A. Stewart et E. Kim, « Amazon's Internal Documents Warn Employees Not to Use Generative AI Models for Work », *Business Insider*, 22 février 2024, disponible sur : www.businessinsider.com.

134. J. Weatherbed, « Midjourney Bans All Stability AI Employees Over Alleged Data Scraping », *The Verge*, 11 mars 2024, disponible sur : www.theverge.com.

135. Ainsi de Cisco et Nvidia, au bénéfice du secteur privé, de Microsoft et Semafor, au profit du secteur journalistique. Lire I. King, « Nvidia, Cisco to Help Companies Build In-House AI Computing », *Bloomberg*, 6 février 2024, disponible sur : www.bloomberg.com ; « Microsoft Partners with Semafor for AI-assisted News Content », Reuters, 5 février 2024, disponible sur : www.reuters.com.

136. À l'image des programmes de Google ou d'Amazon, qui sous couvert de formation de nouveaux « talents », entendent ainsi renforcer les usages. Lire M. Coulter, « Google Pledges 25 Million Euros to Boost AI Skills in Europe », Reuters, 12 février 2024, disponible sur : www.reuters.com ; S. Herrera et C. Cutter, « Amazon Launches Free AI Classes in Bid to Win Talent Arms Race », *Wall Street Journal*, 20 novembre 2023, disponible sur : www.wsj.com.

137. « FTC Launches Inquiry into Generative AI Investments and Partnerships », Federal Trade Commission, 25 janvier 2024, disponible sur : www.ftc.gov.

Une course politique et géopolitique

À certains égards, l'IA semble agiter des clivages politiques internes préexistants entre progressistes et conservateurs. Dans un contexte de polarisation extrême de la scène intérieure américaine, le développement de l'IA générative n'échappe pas aux logiques de tribalisme politique et de « guerre culturelle » à l'œuvre aux États-Unis, qui touchait déjà les réseaux sociaux[138]. ChatGPT a ainsi été rebaptisé « WokeGPT » par ses détracteurs qui estiment que le modèle est orienté politiquement, et Gemini a subi les mêmes critiques. D'où les velléités d'Elon Musk de faire émerger une IA défendant spécifiquement les valeurs conservatrices et libertariennes, supposée contrebalancer l'influence des modèles « progressistes » d'OpenAI, Google, Meta et consorts[139], sans toutefois parvenir à ses fins[140]. Cette tendance n'est pas propre aux seuls États-Unis : au Japon, l'IA est également mobilisée par les conservateurs pour soutenir leur projet politique nationaliste, xénophobe et sexiste[141].

Sur la scène internationale, la course à l'IA a été régulièrement comparée à une nouvelle course aux armements[142], perception accentuée non seulement par certaines déclarations volontairement tonitruantes[143], mais aussi par le durcissement de la rivalité sino-américaine. Ce récit qui s'identifie volontairement à un précédent historique s'est rapidement imposé, alors que la Chine affirme depuis 2017 son ambition de devenir le leader international de l'IA en 2030, et que les États-Unis cherchent à maintenir leur hégémonie technologique par tous les moyens. À Washington, le puissant narratif de l'angoisse du rattrapage voire du dépassement technologique chinois – porté par des faucons de la politique américaine vis-à-vis de la Chine (*China hawks*) comme Eric Schmidt (ancien directeur général de Google) – a notamment contribué à un renforcement de la politique de contrôle des exportations menée par les administrations

138. A. Piquard, « L'intelligence artificielle a déjà les mêmes problèmes que les réseaux sociaux », *Le Monde*, 23 mars 2023, disponible sur : www.lemonde.fr.
139. B. Schreckinger, « Elon Musk's Liberal-trolling AI Plan Has a Core Audience », *Politico*, 17 juillet 2023, disponible sur : www.politico.com.
140. Le *chatbot* Grok, pour partie entraîné avec les données de X et censé incarner cette alternative, s'est visiblement lui aussi vu taxer de « progressisme » intrinsèque. Lire P. Tassi, « Elon Musk's Grok Twitter AI Is Actually "Woke", Hilarity Ensues », *Forbes*, 10 décembre 2023, disponible sur : www.forbes.com.
141. P. Askenazy, « Intelligence artificielle : "Comment conjuguer conservatisme, nationalisme, voire xénophobie, et technophilie ?" », *Le Monde*, 13 mars 2024, disponible sur : www.lemonde.fr.
142. Parmi de nombreux exemples, lire W. Knight, « The Generative AI Boom Could Fuel a New International Arms Race », *Wired*, 7 septembre 2023, disponible sur : www.wired.com ; A. Chow et B. Perrigo, « The AI Arms Race Is Changing Everything », *Time*, 17 février 2023, disponible sur : time.com.
143. On renvoie régulièrement aux déclarations de Vladimir Poutine pour qui « celui qui deviendra leader dans l'IA sera le maître du monde » ou à celles d'Elon Musk pour qui « l'IA est bien plus dangereuse que l'arme nucléaire » et « causera probablement une troisième guerre mondiale ». Lire « After Putin's AI Comments, Elon Musk Imagines World War III », *The Moscow Times*, 5 septembre 2017, disponible sur : www.themoscowtimes.com.

successives[144]. Ce narratif renforce celui de la course aux armements et contribue de fait à empêcher l'émergence d'une gouvernance mondiale unifiée. Les entreprises américaines du secteur, après s'être longtemps montrées réticentes à restreindre leurs échanges avec la Chine, semblent à présent y trouver davantage leur intérêt, dans la mesure où la compétition sino-américaine fournit un argument de poids en faveur de leur propre sous-régulation et de leur accès aux financements publics[145].

Si la vague de l'IA générative a permis à une Silicon Valley engourdie de retrouver son rang après des années d'inquiétude face à la montée en puissance de concurrents chinois[146], la rivalité sino-américaine n'a pas pour autant baissé en intensité. Considérant que la compétition technologique relève d'un jeu à somme nulle, Pékin et Washington se sont engagés dans un bras de fer stratégique, consistant à priver le camp adverse des ressources indispensables au développement de l'IA, à tenter de lui dérober ses secrets industriels[147], et à abaisser les seuils d'acceptabilité quant aux usages potentiels – en particulier militaires.

Bien que cette rivalité n'empêche ni les scientifiques[148] ni les entreprises des deux bords d'échanger discrètement sur les risques inhérents au déploiement de leurs technologies[149], le contexte géopolitique mondial très tendu rend la coopération internationale difficile sur les enjeux de régulation et de gouvernance. Ce d'autant plus que cette compétition concerne avant tout les puissances établies : la grande majorité des pays n'ont ni les moyens financiers ni le savoir-faire technologique nécessaires pour rivaliser avec celles-ci. Leur accès à l'IA de dernière génération sera dès lors davantage déterminé par leurs relations avec la poignée d'États et d'entreprises leaders

144. Sur ce sujet, lire M. Velliet, « Limiter les investissements technologiques vers la Chine. Initiatives et débats aux États-Unis », *Briefings de l'Ifri*, Ifri, 31 août 2023. Parmi les dernières mesures prises par l'administration américaine pour limiter la croissance chinoise dans le secteur de l'IA, la quasi-interdiction faite à NVIDIA, leader de la production de puces spécialisées, d'exporter ses produits de dernière génération en Chine et la nécessité pour des pays tiers (Arabie saoudite et Émirats arabes unis notamment) d'obtenir des licences préalables pour pouvoir acheter auprès de l'industriel américain. Lire D. Sevastopulo et Q. Liu, « US Tightens Rules on AI Chip Sales to China in Blow to Nvidia », *Financial Times*, 17 octobre 2023, disponible sur : www.ft.com. Il faut toutefois rappeler que si les modèles nécessitent des puces en grand nombre pour pouvoir être entraînés, il leur en faut beaucoup moins pour fonctionner une fois mis sur le marché.

145. C. Kang, « A.I. Leaders Press Advantage with Congress as China Tensions Rise », *The New York Times*, 27 mars 2024, disponible sur : www.nytimes.com.

146. G. Nahon, « Intelligence artificielle générative. Le retour en force de la Silicon Valley », *Le Monde*, 2 juillet 2023, disponible sur : www.lemonde.fr.

147. Une affaire récente chez Google AI a rappelé toute l'acuité du problème pour les entreprises américaines – et plus largement occidentales – que représentent les vols de propriété intellectuelle chinois. Lire E. Dou, « Former Google AI Engineer Charged with Stealing Trade Secrets for China Firm », *Washington Post,* 6 mars 2024, disponible sur : www.washingtonpost.com.

148. M. Broersma, « Chinese and Western Scientists Identify "Red Lines" on AI Risks », *Financial Times,* 18 mars 2024, disponible sur : www.ft.com.

149. M. Murgia, « US Companies and Chinese Experts Engaged in Secret Diplomacy on AI Safety », *Financial Times*, 11 janvier 2024, disponible sur : www.ft.com.

du secteur, ce qui ne manquera pas de renforcer les asymétries existantes, voire d'en générer de nouvelles[150].

À l'heure où la prolifération des modèles s'accélère et où les inquiétudes quant aux usages non raisonnés de l'IA se multiplient, une pression accrue pourrait finir par s'exercer sur le duopole sino-américain pour le conduire à prendre davantage de responsabilités dans la gouvernance internationale de ces technologies, et imposer une canalisation voire une sortie par le haut de cette rivalité stratégique confinant à l'impasse. À l'initiative des États-Unis, l'Organisation des Nationes unies a ainsi adopté fin mars sa première résolution appelant à une régulation internationale de l'IA et à mettre cette dernière au service du développement et de la protection des droits humains. Pour autant, il faut souligner que la persistance de divergences Nord-Sud quant à la perception des priorités limite de fait l'établissement d'une approche commune : quand les gouvernements occidentaux cherchent avant tout à établir des règles du jeu au bénéficie des entreprises situées sur leur territoire, ceux de l'Inde, de l'Amérique du Sud et d'autres régions en développement craignent que ces initiatives n'entravent, chemin faisant, leur propre développement économique. La gouvernance internationale de l'IA, partout brandie comme une nécessité, risquerait ainsi de devenir elle-même un narratif, pour mieux se voir vidée de son sens.

Une course à la souveraineté

Face à la force de frappe financière des États-Unis et de la Chine, la peur du « déclassement économique » constitue pour bon nombre d'États un moteur d'investissement et de lobbying politique en faveur de l'IA – y compris en France[151]. Les annonces d'investissements publics constituent dans ce cadre le prolongement de politiques de puissance à l'échelle nationale. Mais la volonté de ne pas abandonner le marché prometteur de l'IA générative aux seules entreprises américaines ou chinoises se double aussi d'un impératif de souveraineté technologique. Celui-ci, très présent en France[152], a aussi cours en Allemagne, au Royaume-Uni, aux Émirats arabes unis, en Arabie saoudite

150. P. Scharre, « The Perilous Coming Age of AI Warfare: How to Limit the Threat of Autonomous Weapons », *Foreign Affairs*, 29 février 2024, disponible sur : www.foreignaffairs.com.

151. *A contrario* les espoirs de retombées économiques colossales vont bon train, le rapport de la commission nationale sur l'IA estimant ainsi que le PIB français pourrait connaître une hausse de 250 à 420 milliards d'euros d'ici 2034 en fonction du degré d'adoption de ces technologies. Lire A. Piquard, « Intelligence artificielle : un plan d'action pour placer la France "à la pointe" », *Le Monde*, 13 mars 2024, disponible sur : www.lemonde.fr.

152. Emmanuel Macron évoque régulièrement ces enjeux et a été jusqu'à parler de « génie français » pour promouvoir les *start-ups* nationales du secteur (Hugging Face, Mistral AI, LightOn) ou de « défi civilisationnel » au sujet de la nécessité de développer des modèles en langue nationale. Lire A. Piquard, « Intelligence artificielle : "La France est loin d'être la seule à penser qu'il s'agit d'un enjeu de souveraineté" », *Le Monde*, 11 janvier 2024, disponible sur : www.lemonde.fr. Les enjeux de localisation de la *start-up* Poolside (à Paris plutôt qu'à San Francisco), après un tour de financement à capitaux français majoritaires, traduit également ces velléités d'affirmation de souveraineté. Lire J. Marin, « La start-up américaine d'IA Poolside se relocalise à Paris », *L'Usine digitale*, 25 août 2023, disponible sur : www.usine-digitale.fr.

ou en Inde. Chaque fois, elle tient au fait que ces États peuvent se targuer de l'existence d'entreprises nationales développant des modèles d'IA générative alternatifs[153]. La promotion de ces derniers comme des leviers de souveraineté relève d'une double stratégie : marketing pour les entreprises nationales qui se positionnent sur le sujet et géopolitique pour les États qui souhaitent apparaître comme des leaders mondiaux, entérinant ainsi l'acception de l'IA comme objet de puissance.

Dans la mesure où l'immense majorité des corpus d'entraînement des LLM sont rédigés en anglais, la question de la langue d'entraînement des modèles fait également l'objet de débats sur la souveraineté, alors que de Paris[154] à Singapour[155], en passant par Johannesburg[156], les appels et les initiatives pour une plus grande diversité linguistique se multiplient[157]. L'ensemble de ces velléités poussent dès lors à une forme de « nationalisme de l'IA » (*AI nationalism*) exacerbant la concurrence étatique[158]. Articulé autour de l'émergence de champions nationaux, il tend à la dispersion et la multiplication des investissements et des aventures technologiques – notamment au sein d'une UE qui tente dans le même temps d'afficher un visage d'unité sur ces enjeux.

Cette course à la souveraineté trouve sa déclinaison dans le champ de la régulation, qui apparaît – notamment pour les États souhaitant échapper au duopole sino-américain – comme un vecteur d'émancipation et d'affirmation de leurs ambitions d'autonomie. Mais plus qu'une véritable compétition normative, la multiplication des efforts de régulation au niveau international forme un « complexe de régimes » interconnectés[159]. Elle traduit avant tout une nécessité pour les États et les organisations internationales d'apparaître comme acteurs et non simples spectateurs du changement qui s'opère sous leurs yeux. Cela peut notamment se traduire par des tentatives plus ou moins

153. On peut notamment mentionner MistralAI (et ses LLM Mistral 7B et Mistral Large) en France, Aleph Alpha (Luminous) en Allemagne, Falcon Foundation (Falcon 180-B) à Abou Dabi, Watad (Mulhem) en Arabie saoudite, CoRover (BharatGPT) ou Kutrim AI (Kutrim) en Inde.

154. Ainsi du *data hub* « Villers-Cotterêts », destiné à renforcer la présence du français dans les modèles d'IA. Lire F. Debès, « Intelligence artificielle : la France se lance dans la bataille culturelle des données », *Les Échos*, 12 décembre 2023, disponible sur : www.lesechos.fr.

155. « Biased GPT? Singapore Builds AI Model to "Represent" Southeast Asians », *Rappler*, 10 février 2024, disponible sur : www.rappler.com.

156. A. Tsanni, « This Company Is Building AI for African Langages », *MIT Technology Review*, 17 novembre 2023, disponible sur : www.technologyreview.com.

157. De nombreux LLM en langue locale (odia, tamil, telugu, malayalam, bengali, marathi, kannada, gujarati, etc.) sont ainsi développés en Inde, pays multilingue par excellence. Lire H. Singh, « India's AI Leap: 10 LLMs That Are Built in India », *Analytics Vidhya*, 19 avril 2024, disponible sur : www.analyticsvidhya.com.

158. « Welcome to the Era of AI Nationalism », *The Economist*, 1er janvier 2024, disponible sur : www.economist.com.

159. R. Csernatoni, « Charting the Geopolitics and European Governance of Artificial Intelligence », Carnegie Europe, 6 mars 2024, disponible sur : https://carnegieeurope.eu.

fructueuses d'établir un rapport de force avec les entreprises du secteur, comme observé en Inde[160].

Ces efforts de régulation s'appuient sur divers récits en fonction des espaces considérés. Le narratif américain du jeu à somme nulle dans la compétition engagée avec Pékin tend à rendre critique toute tentative de régulation ambitieuse aux États-Unis. Toutefois, cela n'empêche pas une forme d'« inflation législative » interne, qui traduit les rapports de force intra-nationaux : il s'agit pour les divers pouvoirs en place d'apporter une forme de réponse aux inquiétudes des électeurs et d'affirmer leur rôle de régulateur[161]. En Chine, le récit des « nouvelles forces productives de qualité » fait de l'IA l'un des principaux piliers sur lesquels le régime chinois entend s'appuyer pour relancer la croissance nationale.

Contrairement à ce que pourrait laisser croire la multiplication des règlements dédiés au cours des dernières années, la régulation du secteur – hors censure des contenus en ligne – demeure donc souple, et l'alliance objective entre entreprises d'IA et gouvernement semble vouée à durer[162]. Au sein de l'UE, le narratif qui sous-tend les efforts de régulation est celui de la priorité donnée à l'émergence d'IA responsables, centrées sur l'humain et dignes de confiance[163]. Incarné par l'*AI Act*, il est également l'expression des velléités de souveraineté européenne et d'une potentielle « troisième voie » face au duopole sino-américain, mais n'est pas sans refléter là aussi des dissonances internes[164]. Les pressions multiples subies par la Commission européenne, de la part de lobbys comme d'États, traduisent de fait l'âpreté de ces enjeux de souveraineté[165]. De son côté, le Royaume-Uni a d'abord tenté de jouer sa propre partition en se positionnant à équidistance de l'UE et des États-Unis – sans que cela ne

160. La volonté indienne de soumettre à approbation gouvernementale les produits d'IA avant tout lancement sur le marché national a été malmenée après s'être heurtée aux critiques tous azimuts. Lire M. Singh, « India Drops Plan to Require Approval for AI Model Launches », *Techcrunch*, 15 mars 2024, disponible sur : techcrunch.com.

161. R. Heath, « Exclusive: States Are Introducing 50 AI-related Bills per Week », *Axios*, 14 février 2024, disponible sur : www.axios.com.

162. Z. Yang, « Why the Chinese Government Is Sparing AI from Harsh Regulations – for Now », *MIT Technology Review*, 9 avril 2024, disponible sur : www.technologyreview.com.

163. À ce titre de comparaison, le décret de l'administration Biden insiste sur l'IA « sûre, sécurisée et digne de confiance ». Lire « President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence », Maison-Blanche, 30 octobre 2023, disponible sur : www.whitehouse.gov.

164. Paris comme Berlin se sont ainsi vus reprocher leurs efforts pour atténuer le projet de réglementation lors des négociations. À ce titre, l'accord tripartite entre la France, l'Italie et l'Allemagne signé en amont de l'adoption de l'*AI Act* européen traduisait cette double logique, selon laquelle la régulation apparaît à la fois comme un obstacle potentiel au développement de leurs entreprises nationales et comme un levier d'affirmation de leur souveraineté collective. Lire A. Rinke, « Germany, France and Italy Reach Agreement on Future AI Regulation », Reuters, 20 novembre 2023, disponible sur : www.reuters.com.

165. Notamment de la part d'OpenAI ou du Département d'État américain. Lire B. Perrigo, « Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation », *Time*, 20 juin 2023, disponible sur www.time.com ; P. Martin, J. Deustch et A. Edgerton, « US Warns EU's Landmark AI Policy Will Only Benefit Big Tech », *Bloomberg*, 6 octobre 2023, disponible sur : www.bloomberg.com.

l'exonère des pressions exercées par les géants technologiques américains[166] – avant d'opter lui aussi pour l'option régulatrice[167].

Le positionnement britannique a plutôt trouvé son incarnation dans le Sommet mondial de l'IA qui s'est tenu à Londres en novembre 2023. En accueillant dans la foulée du G7 d'Hiroshima le premier événement à la fois multilatéral et multi-acteurs d'ampleur sur les risques inhérents à l'IA générative, le Royaume-Uni a non seulement obtenu un succès diplomatique, mais aussi créé une dynamique de coopération potentiellement prometteuse[168]. Avant même le sommet, le pays – sous l'impulsion personnelle de Rishi Sunak, qui souhaite en faire son legs politique – avait démontré sa volonté d'institutionnaliser le champ de la régulation en annonçant la création d'un AI Safety Institute. Appelé à être dupliqué dans les autres États parties au sommet (comme c'est déjà le cas aux États-Unis), celui-ci, au-delà de ses attributions initiales, pourrait évoluer pour devenir un pourvoyeur de normes et renforcer le leadership politique britannique sur ces enjeux[169].

Dans le sillage de Londres, Séoul et Paris ont souhaité organiser les prochains points d'étape, afin de solidifier le processus engagé, mais aussi de mettre en valeur leurs écosystèmes nationaux et leur propre approche des enjeux. Tenu en mai 2024, le sommet de Séoul a mis l'accent sur la sécurité des modèles et abouti à un engagement en ce sens de 16 entreprises majeures du secteur (américaines, chinoises, coréennes et émiraties). Il a également entériné la création d'un réseau d'instituts analogues au modèle britannique et l'engagement des États à déterminer divers seuils de risques pour les modèles les plus avancés[170]. Pour la France, qui accueillera le sommet suivant dans quelques mois, l'enjeu sera notamment de le rendre plus inclusif (davantage ouvert à la société civile), d'élargir les thématiques abordées et de maintenir un équilibre entre enjeux de sécurité et soutien à l'entreprenariat. Son organisation pourrait également être l'occasion de proposer des

166. C. Criddle, A. Gross et M. Murgia, « World's Biggest AI Tech Companies Push UK over safety Tests », *Financial Times*, 7 février 2024, disponible sur : www.ft.com. Notons que les grandes entreprises du secteur ne semblent pas vouloir jouer pleinement le jeu des accords volontaires *a posteriori*. Lire V. Manancourt, G. Volpicelli et M. Chatterjee, « Rishi Sunak Promised to Make AI Safe: Big Tech's Not Playing Ball », *Politico*, 26 avril 2024, disponible sur : www.politico.com.

167. Après s'être montré réticent à établir une loi nationale pour réguler le secteur, le Royaume-Uni semble à présent s'y résoudre. Lire D. Mosolova, « UK Will Refrain from Regulating AI "in the Short Term" », *Financial Times*, 16 novembre 2023, disponible sur : www.ft.com ; A. Gross et C. Criddle, « UK Rethinks AI Legislation as Alarm Grows over Potential Risks », *Financial Times*, 15 avril 2024, disponible sur : www.ft.com.

168. Notamment *via* la déclaration de Bletchley, engagement conjoint de 28 gouvernements – dont la Chine et les États-Unis – et de grandes entreprises d'IA appelant à soumettre les modèles d'IA avancée à des contrôles de sécurité préalables à leur déploiement. Lire M.-F. Cuéllar, « The UK AI Safety Summit Opened a New Chapter in AI Diplomacy », Carnegie Endowment for International Peace, 9 novembre 2023, disponible sur : carnegieendowment.org.

169. D. Milmo, « UK's AI Safety Institute "Needs to Set Standards Rather Than Do Testing" », *The Guardian*, 11 février 2024, disponible sur : www.theguardian.com.

170. « Seoul Declaration for Safe Innovative and Inclusive AI: AI Seoul Summit 2024 », Department for Science, Innovation and Technology, disponible sur : www.gov.uk.

innovations institutionnelles, telle la création d'une « Organisation mondiale de l'IA[171] » ou à défaut, d'une Organisation mondiale des données[172], à même d'apporter des réponses collectives à des défis qui le sont tout autant, et à établir un cadre de coopération constructif sur ces sujets.

Une telle proposition nécessiterait cependant d'intenses efforts diplomatiques pour être concrétisée, alors que se multiplient les initiatives étatiques. Car cette politique de sommets, qui offre à la fois des perspectives aux États en matière de *nation branding* et tente de répondre à un besoin urgent de gouvernance, est appelée à être reproduite ailleurs. Les dirigeants des pays du Sud semblent de fait vouloir emprunter le pas de leurs homologues occidentaux ; le Rwanda a ainsi annoncé l'organisation sur le sol africain du premier sommet international sur l'IA dès cette année[173]. Celui-ci fera office de vitrine non négligeable pour Kigali, qui cherche à se positionner comme le leader africain des technologies numériques et de l'IA en particulier.

171. Comme suggéré notamment dans le rapport de la Commission nationale de l'IA. « IA : notre ambition pour la France », Commission de l'Intelligence artificielle, mars 2024.
172. I. Bremmer, « Why We Need a World Data Organization. Now », *GZERO*, 25 novembre 2019, disponible sur : www.gzreomedia.com.
173. « Rwanda Announces Plans to Host Inaugural High-Level Summit on AI in Africa », Rwanda Centre for the Fourth Industrial Revolution, 18 janvier 2024, disponible sur : https://c4ir.rw.

Conclusion

Du fait du potentiel de transformation des sociétés – encore mal cerné – dont l'IA générative est porteuse, sa grande accessibilité génère des craintes légitimes auxquelles il sera nécessaire d'apporter des réponses collectives. Il importe pour cela de pouvoir distinguer ce qui relève des fantasmes ou des fausses promesses d'un côté, et des risques substantiels de l'autre. Mais ces craintes sont aussi le reflet d'intérêts plus ou moins divergents, à l'heure où le développement de ces technologies nourrit des rivalités à de multiples échelles. Le tournant discursif dont l'IA fait l'objet doit ainsi être perpétuellement déconstruit pour mieux appréhender la réalité des risques inhérents au déploiement de ces technologies.

Les discours abondants qui sont produits sur l'IA tendent à noyer les problématiques urgentes dans la masse des risques potentiels ou à venir. À rebours des récits long-termistes faisant diversion, certains enjeux sont bien de nature immédiate et systémique, à l'image des impacts sociaux-économiques inégaux de l'IA sur les diverses sociétés à travers le monde, ou de son poids énergétique et son empreinte carbone exponentiels. La question de la gestion de ces externalités négatives pourrait ressurgir avec d'autant plus de violence dans l'espace public que celles-ci auront été ignorées, voire minorées du fait de récits plus bruyants et plus influents. Il y va possiblement de l'aggravation de clivages sociopolitiques préexistants, dans des démocraties déjà fragilisées et menacées.

Alors que la gouvernance internationale de l'IA n'est encore qu'un vaste chantier, une lecture avisée des antagonismes aussi bien politiques, géopolitiques, économique que sociaux paraît donc impérative pour pouvoir dépasser le paravent des discours et espérer orienter la régulation dans le sens du bien commun. Il appartiendra aux régulateurs d'opérer les distinctions nécessaires et d'œuvrer à une éventuelle convergence réglementaire qui, pour le moment, demeure une perspective lointaine tant la course à l'IA se pense comme étant féroce et inarrêtable.

27 rue de la Procession 75740 Paris cedex 15 – France

Ifri.org