



INSTITUT DE LA
SOUVERAINETÉ
NUMÉRIQUE



Intelligence Artificielle

Enjeux et perspectives pour les
droits humains en Europe

Dirigé par Bernard Benhamou
Secrétaire général de l'ISN

TABLE DES MATIÈRES

| | |
|--|-----------|
| I. PRÉSENTATION ET RÉSUMÉ ANALYTIQUE | 5 |
| II. INTRODUCTION | 8 |
| III. L'INTELLIGENCE ARTIFICIELLE LEVIER DU REMODELAGE DES SOCIÉTÉS | 12 |
| A. Données personnelles, profils d'utilisateurs et microciblage | 12 |
| B. Une mutation dans la transmission des savoirs | 16 |
| C. Régulation européenne de l'IA : des objectifs industriels et politiques | 19 |
| IV. DES LIBERTÉS FONDAMENTALES EN PÉRIL ? | 24 |
| A. IA & désinformation : les nouvelles menaces sur la démocratie | 26 |
| B. Des formes inédites de manipulations politiques | 31 |
| C. Vers des relations affectives avec les IA ? | 33 |
| V. LES CONSÉQUENCES DES IA SUR LES DÉMOCRATIES | 35 |
| A. Des IA protectrices contre des IA hostiles ? | 35 |
| B. Les nouvelles formes de lutte contre la désinformation | 38 |
| C. Une course au moins-disant démocratique ? | 40 |
| D. Les vulnérabilités des démocraties face à l'IA | 43 |
| E. Chine : l'IA au service du contrôle politique des populations | 45 |

| | |
|---|-----------|
| VI. TRANSHUMANISME, LONG-TERMISME, EUGÉNISME : DES IDÉOLOGIES AU SERVICE D'UNE VISION RADICALE DE L'HUMAIN | 51 |
| A. Transhumanisme : vers des humains inhumains ? | 52 |
| B. IA et transformations culturelles des sociétés | 56 |
| C. Du transhumanisme à l'eugénisme | 58 |
| VII. PERSPECTIVES TECHNOLOGIQUES DES IA ET ÉVOLUTIONS POLITIQUES | 61 |
| A. Des IA encore éloignées de « l'intelligence artificielle générale » | 61 |
| B. De la science-fiction à la réalité industrielle | 64 |
| C. IA, données et modèles économiques toxiques | 68 |
| VIII. IA ET SANTÉ : STRATÉGIES ET ENJEUX POLITIQUES | 71 |
| A. Données de santé : perspectives industrielles et politiques | 71 |
| B. Modifier le centre de gravité de l'économie de la santé | 75 |
| C. Quelle stratégie pour les données de santé en Europe ? | 77 |
| IX. ARCHITECTURE ET GOUVERNANCE DE L'IA | 81 |
| X. IA ET PROPRIÉTÉ INTELLECTUELLE : NOUVELLES FRONTIÈRES JURIDIQUES | 86 |
| A. Licences et régulation du code des IA | 86 |
| B. IA : quel impact sur les auteurs et les créateurs ? | 88 |
| C. IA génératives : les nouvelles frontières du droit | 90 |
| XI. LES DÉFIS ÉTHIQUES LIÉS AUX IA | 92 |
| A. Quelles valeurs humaines et politiques seront transmises aux AI ? | 92 |
| B. Les difficultés de « l'alignement » des IA | 93 |

| | | |
|--------------|---|------------|
| C. | La confiance, objectif stratégique pour les concepteurs des IA | 95 |
| D. | Des IA étroitement intégrées à la vie sociale, culturelle et politique | 98 |
| XII. | LES ENJEUX GÉOPOLITIQUES DE L'IA | 102 |
| A. | Des guerres pour l'IA aux guerres via l'IA | 103 |
| B. | États-Unis : ralentir mais pas étouffer les technologies d'IA chinoises | 104 |
| C. | Des IA de plus en plus énergivores et pourtant cruciales pour le développement durable | 107 |
| D. | Fin de la globalisation : vers le « friendshoring » ? | 109 |
| E. | Affrontement entre modèles politiques et technologiques américains, chinois... et européens | 112 |
| F. | L'Union européenne : citadelle démocratique assiégée par trois empires ? | 114 |
| G. | Des IA encore plus propices à la surveillance de masse | 116 |
| H. | L'IA au cœur du conflit entre la Chine et Taïwan | 119 |
| XIII. | POLITIQUE INDUSTRIELLE DE L'IA : LES LEVIERS DE LA PROTECTION DES DROITS DE L'HOMME | 123 |
| A. | Vers une « troisième voie » éthique pour les IA européennes | 123 |
| B. | Quelle réponse européenne face à la techno (ou data) colonisation ? | 126 |
| C. | Une synergie nécessaire entre politique industrielle et régulation des technologies | 132 |
| D. | De nouvelles stratégies de politique industrielle pour l'IA | 137 |
| XIV. | RECOMMANDATIONS | 140 |
| XV. | PRÉSENTATION DE L'ISN ET IDFRIGHTS | 143 |
| XVI. | REMERCIEMENTS | 146 |
| XVII. | ORGANISATIONS PARTENAIRES | 148 |

INTELLIGENCE ARTIFICIELLE

Enjeux & perspectives pour les droits humains en Europe

I. PRÉSENTATION ET RÉSUMÉ ANALYTIQUE

Ce rapport présente une synthèse des travaux consacrés à l'intelligence artificielle menés au sein de l'Institut de la Souveraineté Numérique et de l'Institut IDFrighs. Il vise à analyser les évolutions économiques, politiques et culturelles associées au développement de ces technologies, ainsi que leurs effets sur les sociétés et les démocraties européennes. Il est conçu comme un outil de réflexion et de débat à destination des citoyens et des acteurs publics. Ce rapport propose aussi des recommandations afin de favoriser l'émergence d'une « *troisième voie* » européenne en matière d'intelligence artificielle, fondée sur des principes de transparence, de démocratie et de respect des libertés fondamentales.

Les technologies d'intelligence artificielle démultiplient les capacités d'analyse et de traitement des données, facilitent l'aide à la décision et permettent l'automatisation de tâches complexes longtemps considérées comme l'apanage des capacités humaines. Elles trouvent aujourd'hui des applications dans un nombre croissant de secteurs, allant de la gestion de la consommation énergétique au diagnostic précoce de pathologies, en passant par l'optimisation des processus industriels, le pilotage de véhicules autonomes ou encore le développement d'armes « intelligentes ». Parmi ces technologies d'IA, les grands modèles de langage (*Large Language Models*, ou *LLM*) ont connu une expansion d'une rapidité sans précédent. Au début

de l'année 2026, le nombre d'utilisateurs de ces modèles est estimé à près d'un milliard à l'échelle mondiale¹.

Les débats relatifs à l'intelligence artificielle se sont majoritairement concentrés sur ses impacts économiques et sociaux, ou sur des perspectives hypothétiques et lointaines d'une suprématie de systèmes devenus plus « intelligents » que leurs concepteurs. Cette focalisation tend à occulter les conséquences plus immédiates que ces technologies sont susceptibles d'avoir, sur les sociétés européennes et leurs régimes politiques. En raison de l'influence croissante qu'elles exercent sur les libertés individuelles et sur l'évolution des modes de vie, les IA soulèvent des enjeux inédits pour les démocraties européennes. À la différence des innovations technologiques antérieures, dont les effets étaient principalement économiques et industriels, l'IA engage des transformations profondes des rapports sociaux et politiques. Comme le souligne la spécialiste des enjeux sociopolitiques de l'IA Kate Crawford, « *L'intelligence artificielle, c'est la politique par d'autres moyens...* »².

En l'absence de réponses politiques et industrielles coordonnées à l'échelle européenne, ces technologies pourraient constituer de nouvelles menaces pour les libertés publiques, notamment à travers la généralisation de dispositifs de surveillance algorithmique et de manipulation de masse des comportements. Le modèle européen de protection sociale pourrait également être fragilisé par l'émergence de logiques de contrôle permanent des individus fondées sur l'analyse automatisée des données. Par ailleurs, l'essor récent des intelligences artificielles génératives est susceptible de transformer en profondeur la création culturelle et la transmission des connaissances, tout en remettant en question les fondements traditionnels de la propriété intellectuelle. Au-delà de ces dimensions culturelles, l'IA est appelée à intervenir dans les processus démocratiques eux-mêmes, en redéfinissant les conditions du débat public et, potentiellement, la notion de vérité sur laquelle reposent les décisions collectives. En outre, ces technologies contribuent à la « démocratisation » de nouvelles formes d'ingérence dans les processus démocratiques européens, comme l'avait

¹ Digital 2026: More than One Billion People Use AI (Data Reportal 15 Oct 2025)
datareportal.com/reports/digital-2026-one-billion-people-using-ai

² Kate, Crawford. Contre-atlas de l'Intelligence artificielle : Les coûts politiques, sociaux et environnementaux de l'IA (p. 32). Zulma 2022

illustré l'affaire *Cambridge Analytica* aux États-Unis en 2018, ou plus récemment les ingérences russes en Roumanie via la plateforme chinoise *TikTok*.

Face à ces risques, les acteurs technologiques reconnaissent désormais la nécessité d'un encadrement du développement de l'intelligence artificielle afin de prévenir les dérives autoritaires, les discriminations et les nouvelles formes d'ingérence politique. Dans cette perspective, les initiatives européennes, notamment le règlement sur l'intelligence artificielle (*AI Act*), constituent des avancées pionnières. Elles sont toutefois aujourd'hui contestées par les autorités américaines, qui les perçoivent à la fois comme une entrave au développement de leurs principaux acteurs technologiques et comme un obstacle politique à la diffusion de leurs orientations idéologiques en Europe. Néanmoins, la régulation, à elle seule, ne saurait suffire à contenir les risques associés à ces technologies. Dans un contexte de forte dépendance de l'Europe à l'égard de solutions extra-européennes, l'un des défis majeurs pour l'avenir de l'Union européenne réside dans sa capacité à faire émerger des acteurs technologiques de dimension mondiale. Comme le rappelle Mario Draghi dans son rapport sur la productivité européenne³, l'incapacité de l'Europe à mettre en place des politiques de financement ambitieuses pour ces technologies stratégiques pourrait conduire à une forme d'agonie industrielle, économique et politique face à des acteurs extra-européens qui ne partagent ni ses principes ni ses valeurs démocratiques. Ces enjeux prennent une acuité particulière avec la montée des tensions géopolitiques entre les États-Unis et la Chine. L'intelligence artificielle constitue en effet un élément central de la suprématie économique et militaire, mais également l'instrument privilégié des nouvelles formes des guerres hybrides qui visent les opinions publiques européennes. Ainsi, en l'absence d'acteurs industriels européens capables de constituer des alternatives aux technologies américaines ou chinoises, l'Europe pourrait devenir durablement une colonie technologique et politique de deux autres continents.

³The future of European competitiveness – A competitiveness strategy for Europe (Report by Mario Draghi Sept 2024) https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

II. INTRODUCTION

Les technologies d'intelligence artificielle (IA) ont aujourd'hui d'innombrables champs d'application, allant de la gestion de la consommation d'énergie au diagnostic précoce des pathologies, en passant par l'optimisation des processus industriels, le pilotage des véhicules autonomes, ou encore la conception d'armes intelligentes. Le développement des technologies d'intelligence artificielle a ainsi permis de démultiplier les capacités de traitement des données, facilitant l'aide à la décision et l'automatisation de tâches complexes que l'on croyait durablement réservées aux humains. Pour les industriels du secteur, l'IA représente à la fois un enjeu technologique majeur et une opportunité d'établir une nouvelle cartographie de la puissance industrielle, financière et politique. Ainsi, il y a huit ans, bien avant l'essor inédit des IA génératives (comme *ChatGPT*, *Gemini*, *Perplexity*, *Claude* ou *Mistral*), Sundar Pichai, PDG de *Google*, décrivait le futur des technologies en ces termes : « *La prochaine grande étape des technologies correspondra à la disparition des appareils eux-mêmes. Au fil du temps, l'ordinateur (quelle que soit sa forme) deviendra un assistant intelligent qui vous aidera tout au long de votre journée. Nous allons passer de l'ère du mobile à l'ère de l'intelligence artificielle...⁴* ».

Plus récemment, Mustafa Suleyman, le cofondateur de la société *DeepMind* (devenue filiale de *Google*), décrivait en ces termes le rôle central que les IA occuperont dans les temps à venir : « *L'IA n'est pas une simple technologie de plus ; elle est bien plus profonde et plus puissante que cela. Le risque réside moins dans sa surmédiation que dans l'échec à mesurer l'ampleur de la vague à venir. L'IA n'est pas seulement un outil ou une plateforme, c'est une métatechnologie transformatrice, la technologie derrière la technologie et tout le reste, elle-même*

⁴ "Looking to the future, the next big step will be for the very concept of the "device" to fade away. Over time, the computer itself—whatever its form factor—will be an intelligent assistant helping you through your day. We will move from mobile first to an AI first world."

Sundar Pichai PDG de Google, Lettre des fondateurs du 28 avril 2016
<https://googleblog.blogspot.fr/2016/04/this-years-founders-letter.html>

productrice d'outils et de plateformes ; ce n'est pas seulement un système, c'est un générateur de systèmes de toutes natures⁵. »

Avec l'essor des IA génératives⁶, ces technologies pourraient aussi transformer le rapport de nos sociétés à la création culturelle et à la transmission des

**“ L'intelligence
artificielle, c'est la
politique par
d'autres moyens...”**

Kate Crawford

connaissances, modifiant ainsi en profondeur le fonctionnement même de nos sociétés. Au-delà de leurs effets industriels et sociaux, les technologies d'IA pourraient avoir des conséquences durables sur l'organisation de nos sociétés et sur nos libertés. En effet, qu'il s'agisse de l'organisation des processus démocratiques, des conditions d'exercice du débat public ou de la compréhension même de la notion de vérité, ces

technologies pourraient transformer l'ensemble des fonctions et des activités politiques. Ainsi, pour Kate Crawford, spécialiste des enjeux de société de l'IA, l'étendue des champs d'application actuels et futurs de l'IA confère aux acteurs des technologies un rôle politique majeur. Pour appuyer sa démonstration, Kate Crawford paraphrase même les mots célèbres de Clausewitz sur la guerre, mais cette fois, c'est l'intelligence artificielle qui est décrite comme la continuation de la politique par d'autres moyens.

⁵ AI is far deeper and more powerful than just another technology. The risk isn't in overhyping it; it's rather in missing the magnitude of the coming wave. It's not just a tool or platform but a transformative meta-technology, the technology behind technology and everything else, itself a maker of tools and platforms, not just a system but a generator of systems of any and all kinds.

Suleyman, Mustafa; Bhaskar, Michael. La Déferlante : Technologie, pouvoir et le dilemme majeur du XXI^e siècle (p. 95). Fayard 2023; Ouvrage original : Suleyman, Mustafa; Bhaskar, Michael. The Coming Wave (p. 78). Random House 2023

⁶ En particulier les « grands modèles de langage » (LLM ou Large Language Models) comme ChatGPT d'OpenAI ou Gemini de Google, mais aussi les IA destinées à la création d'images ou de vidéos à partir de textes comme Midjourney, Stable Diffusion, DALL·E 3, etc.

« L'intelligence artificielle est désormais un acteur qui détermine ce qu'est le savoir, la communication ou le pouvoir. Cette reconfiguration se produit au niveau de l'épistémologie, des principes de justice, de l'organisation sociale, de l'expression politique, de la culture, de la compréhension des corps humains, des subjectivités et des identités : ce que nous sommes et ce que nous pouvons être. Mais allons plus loin. Parce qu'elle recartographie le monde et y intervient, l'intelligence artificielle, c'est la politique par d'autres moyens, même si on la reconnaît rarement comme telle. Cette politique est menée par les grandes maisons de l'IA, soit la demi-douzaine d'entreprises qui dominent l'informatique planétaire⁷. »

Pour l'historien Yuval Harari, les transformations rendues possibles par l'IA méritent d'être analysées et contrôlées pour éviter que les sociétés démocratiques ne sombrent dans des dérives autocratiques en raison d'un usage incontrôlé de ces technologies : *« Les technologies qui ont favorisé la démocratie sont en train de changer et, à mesure que l'intelligence artificielle se développe, elles pourraient encore évoluer. Les technologies de l'information continuent de progresser, et les biotechnologies commencent à nous ouvrir une fenêtre sur notre vie intérieure – nos émotions, nos pensées et nos choix. Ensemble, les technologies de l'information et les biotechnologies créeront des bouleversements sans précédent dans la société humaine, en érodant l'action humaine et, peut-être, subvertiront-elles aussi les désirs humains. Dans de telles conditions, la démocratie libérale et l'économie de marché pourraient devenir obsolètes⁸. »*

⁷ Simply put, artificial intelligence is now a player in the shaping of knowledge, communication, and power. These reconfigurations are occurring at the level of epistemology, principles of justice, social organization, political expression, culture, understandings of human bodies, subjectivities, and identities: what we are and what we can be. But we can go further. Artificial intelligence, in the process of remapping and intervening in the world, is politics by other means—although rarely acknowledged as such. These politics are driven by the Great Houses of AI, which consist of the half-dozen or so companies that dominate large-scale planetary computation.

Kate, Crawford. Contre-atlas de l'Intelligence artificielle : Les coûts politiques, sociaux et environnementaux de l'IA (p. 32). Zulma 2022. Ouvrage original : Crawford, Kate. Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence (pp. 19-20). Yale University Press 2021

⁸ The technology that favored democracy is changing, and as artificial intelligence develops, it might change further. Information technology is continuing to leap forward; biotechnology is beginning to provide a window into our inner lives—our emotions, thoughts, and choices. Together, infotech and biotech will create

Les défis liés à l'intégration des préoccupations humaines dans la conception des IA relèvent désormais d'enjeux sociaux et politiques majeurs. Ces défis seront d'autant plus importants que certains acteurs technologiques commencent désormais à déployer un « agenda politique » au travers de leurs IA. Le développement de ces technologies en Europe devra s'accompagner d'une réflexion sur la nature même de nos interactions sociales et politiques, et sur la manière dont nous pourrions, ou non, permettre que certaines de ces interactions soient modifiées par les acteurs des technologies. Cette adéquation entre des préoccupations éthiques et politiques et la réalisation des IA est ce que Brian Christian, auteur d'un ouvrage de référence sur l'éthique de l'IA, nomme le « *problème d'alignement de l'IA* » : « *L'apprentissage automatique est un domaine manifestement technique qui se heurte de plus en plus à des problématiques humaines. Nos dilemmes humains, sociaux et civiques deviennent techniques. Et nos dilemmes techniques deviennent humains, sociaux et civiques. Nos succès et nos échecs à faire en sorte que ces systèmes fassent « ce que nous voulons » nous offrent un miroir à la fois révélateur et sans complaisance⁹* ». Ainsi, parallèlement aux mesures qui permettront de développer en Europe les prochaines générations d'IA, l'un des principaux défis politiques et industriels auxquels nos sociétés seront confrontées sera de veiller à ce que ces technologies restent le reflet des principes de liberté et des valeurs démocratiques européennes.

unprecedented upheavals in human society, eroding human agency and, possibly, subverting human desires. Under such conditions, liberal democracy and free-market economics might become obsolete.

Why Technology Favors Tyranny (The Atlantic, Oct 2018)

www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/

⁹ « Machine learning is an ostensibly technical field crashing increasingly on human questions. Our human, social, and civic dilemmas are becoming technical. And our technical dilemmas are becoming human, social, and civic. Our successes and failures alike in getting these systems to do “what we want,” it turns out, offer us an unflinching, revelatory mirror. »

Brian Christian. The Alignment Problem (W. W. Norton & Company 2020)

III. L'INTELLIGENCE ARTIFICIELLE LEVIER DU REMODELAGE DES SOCIÉTÉS

A. Données personnelles, profils d'utilisateurs et microciblage

Dès les premières phases du développement de grandes plateformes de l'Internet, un processus d'extraction des données des utilisateurs a été mis en place. Cela a été en particulier le cas pour *Google* avec ce que Shoshana Zuboff nomme le « surplus comportemental » (ou *behavioral surplus*) dans son ouvrage *L'Âge du capitalisme de surveillance*¹⁰. Ce procédé a été inventé par *Google* à partir de l'analyse des requêtes des utilisateurs qui n'étaient pas utiles à l'amélioration du moteur de recherche. En effet, dans un premier temps, *Google* n'archivait que les requêtes pertinentes pour l'optimisation des résultats du moteur de recherche. Par la suite, la totalité des requêtes a été archivée pour créer des profils de l'ensemble des utilisateurs. Ce « surplus » d'informations a permis à *Google* d'analyser massivement les comportements de plusieurs centaines de millions puis de milliards de personnes. Ces informations permettaient à *Google* d'offrir aux annonceurs publicitaires un ciblage précis en fonction du profil de chaque utilisateur reposant sur plusieurs milliers de paramètres sur chaque individu. On parle alors de microciblage (ou *microtargeting*). Ces évolutions ont permis à *Google* de devenir la plus grande régie publicitaire au monde. Certains économistes présentent même *Google* comme « une régie publicitaire qui possède un moteur de recherche... »

Mais, au-delà de leur utilisation publicitaire, ces données de microciblage peuvent avoir des utilisations à la fois sociales et politiques (comme dans le cas de l'affaire *Cambridge Analytica/Facebook*). Ainsi, comme le notent Zuboff et Harari, à l'issue de la constitution des profils de ses utilisateurs, *Google* devenait capable d'analyser

¹⁰ Shoshana Zuboff. *The Age of Surveillance Capitalism* (Profile Books, Jan 2019)

l'évolution des comportements individuels, mais aussi leurs conséquences sur des groupes spécifiques ou sur l'ensemble d'une population. Ce qui faisait dire à

“*Nous pensions chercher sur Google, mais en fait c’est Google qui cherchait en nous...*”

Shoshana Zuboff

Shoshana Zuboff : « *Nous pensions chercher sur Google, mais en fait c’est Google qui cherchait en nous...* ». Pour Madhumita Murgia, spécialiste de l'IA au *Financial Times*, au-delà des moteurs de recherche, chaque action des usagers participe désormais à la constitution de ces profils ultra-détaillés. Ces profils qui seront par la suite utilisés par des IA pour modéliser les comportements ou interagir avec les usagers dans le cadre de nouvelles générations de services à valeur ajoutée.

J’ai commencé à détricoter la structure de cette florissante économie des données. Chaque fois que j’interagissais avec un produit en ligne – par exemple Google Maps, Uber, Instagram ou des cartes de crédit sans contact – d’un simple clic, mon comportement était enregistré par ces petits cookies. Associés à des informations publiques telles que mon impôt foncier ou mon dossier électoral, ainsi qu’à mes habitudes d’achat en ligne et à des informations de localisation en temps réel, ces ensembles de données inoffensives pouvaient révéler beaucoup de choses sur moi, de mon sexe et de mon âge jusqu’aux nuances de ma personnalité et de mes prises de décisions futures¹¹.

Il convient donc pour *Google* et plus largement pour l'ensemble des acteurs dont le modèle économique repose sur la publicité, de maintenir ce canal d'extraction de données issues des usagers. La « philosophie » de l’hypertransparence des

¹¹ I started to unpick the structure of this flourishing data economy. Every time I interacted with an online product – say Google Maps, Uber, Instagram, or contactless credit cards – with a single click, my behaviour was logged by these little cookies. Combined with public information such as my council tax or voter records, along with my online shopping habits and real-time location information, these benign datasets could reveal a lot about me, from my gender and age, down to nuances about my personality and my future decision-making. Madhumita Murgia: Code Dependent: Living in the Shadow of AI (Henry Holt and Co. 2024)

individus est ainsi devenue le cœur du modèle économique des grandes plateformes publicitaires de l'Internet. Cette évolution correspondait aussi à la fin programmée de la notion de vie privée. Ce qu'Éric Schmidt, alors PDG de *Google*, résumait ainsi : « *Si vous faites une chose que vous voulez que personne ne sache, peut-être devriez-vous déjà commencer par ne pas la faire*¹² ». Il convient en effet pour ces plateformes de ne jamais tarir ce flux d'informations sur les utilisateurs qui leur permettent de monétiser leurs services. Les acteurs technologiques sont ainsi passés de modèles économiques « productivistes » à des modèles « extractivistes » où l'objectif est d'extraire le plus de données possible sur leurs usagers. Cette extraction des données et contenus issus des utilisateurs de l'Internet est devenue d'autant plus importante que ces données participent désormais à l'entraînement des modèles d'IA. De plus, l'analyse des données des utilisateurs via des technologies d'IA est ainsi devenue une activité majeure des acteurs publicitaires de l'Internet comme *Google* ou *Facebook*. Ainsi, dans un article au titre évocateur, « *Ce que 7 brevets effrayants révèlent sur Facebook*¹³ », le *New York Times* décrivait dès 2018 les méthodes utilisées par Facebook pour analyser dans le détail le profil psychologique et le comportement de ses usagers. Ces brevets portaient sur l'analyse de leurs habitudes de consommation, leurs déplacements en temps réel, les évolutions de leurs liens avec les autres usagers, leur consultation de contenus écrits ou audiovisuels ou encore l'identification des auteurs de photos et vidéos postées sur Internet à partir d'infimes défauts présents dans les capteurs de leurs smartphones.

Ces profils ultra-détaillés que rassemblent les courtiers en données (ou *data brokers*) permettent ainsi à des acteurs publicitaires, mais aussi à des États, des groupes politiques ou religieux d'élaborer des contenus ciblés pour influencer à chaque instant les utilisateurs. Cela permet de faire évoluer leurs habitudes de consommations mais aussi leurs convictions idéologiques ou religieuses. Et comme le précisait la représentante américaine Kathy Castor lors de l'audition de

¹² Google CEO Eric Schmidt Dismisses the Importance of Privacy (Electronic Frontier Foundation, 10 Dec 2009) www.eff.org/fr/deeplinks/2009/12/google-ceo-eric-schmidt-dismisses-privacy.

¹³ What 7 Creepy Patents Reveal about Facebook (New York Times, 21 juin 2018) www.nytimes.com/interactive/2018/06/21/opinion/sunday/facebook-patents-privacy.html

Mark Zuckerberg au *Congrès des États-Unis*¹⁴, la constitution par *Facebook* de ces profils extrêmement détaillés ne porte pas uniquement sur les utilisateurs de *Facebook* mais aussi sur des personnes qui ne sont pas enregistrées sur la plateforme et qui visitent d'autres sites web. On évoque alors le terme de « *profils fantômes* » (*shadow profiles*) qui permettent à ces sociétés de consolider et d'étendre leurs bases de profils à la quasi-totalité des utilisateurs de l'Internet.

Ainsi, parallèlement aux données indexées par le moteur de recherche, ces données sur les utilisateurs se révèlent être un instrument crucial pour entraîner des modèles d'IA qui peuvent désormais prétendre remplacer les moteurs de recherche traditionnels par des « *moteurs de réponses* » (ou *answer engines*)¹⁵.

Par la facilité avec laquelle ils permettent d'accéder à des données qui nécessitaient auparavant un effort de mémoire individuelle mais aussi collective, les smartphones ont souvent été décrits comme notre « exo-cerveau ». Avec la montée en puissance des assistants intelligents, ce sont des fonctions d'analyse et de traitement des informations qui pourraient être transférées à des IA. Ainsi, en proposant leurs substituts aux moteurs de recherche d'informations traditionnels, des sociétés comme *OpenAI* (avec *SearchGPT*), ou encore *Perplexity AI*, souhaitent démontrer que les fonctions des moteurs comme *Google* peuvent être remplacées par des modèles d'IA génératives¹⁶. Ces nouvelles générations de services offrent en effet des réponses structurées aux questions des utilisateurs et plus seulement des résultats sous forme de listes de liens. À mesure qu'ils sont massivement adoptés, ces « moteurs de réponses » constituent de nouveaux filtres entre les usagers et les informations disponibles sur Internet. Avec, là encore, des risques de dérive et de censure imperceptible ou au contraire d'amplification

¹⁴ U.S. Rep. Castor Questions Mark Zuckerberg, CEO of Facebook (C-Span, 11 Apr 2018)
www.youtube.com/watch?v=OIvq763F57k

¹⁵ AI search could break the web (MIT Tech Review, 31 Oct 2024)
www.technologyreview.com/2024/10/31/1106504/ai-search-could-break-the-web/

¹⁶ OpenAI Is Launching Search Engine, Taking Direct Aim at Google (Wall Street Journal, 25 Jul 2024)
www.wsj.com/tech/ai/openai-search-engine-searchgpt-97771f86

algorithmique de certains contenus. Les questions liées à la transmission des savoirs à l'ère des IA génératives pourraient ainsi devenir centrales sur le plan économique, mais aussi culturel et politique dans les années à venir.

B. Une mutation dans la transmission des savoirs

L'éducation est le moment où nous décidons si nous aimons suffisamment le monde pour en assumer la responsabilité, et du même coup le sauver de cette ruine qui, sans le renouveau, sans l'arrivée des nouveaux et des jeunes, serait inévitable. Et c'est aussi dans l'éducation que nous décidons si nous aimons suffisamment nos enfants pour ne pas les expulser de notre monde et les abandonner à eux-mêmes, ni pour leur ôter la chance d'entreprendre quelque chose de nouveau, quelque chose d'imprévu pour nous, mais pour les préparer à la tâche de renouveler un monde commun.¹⁷

Hannah Arendt
La crise de la Culture (1961)

La transmission des savoirs et connaissances au sein des sociétés humaines a considérablement évolué depuis l'apparition des ordinateurs et l'effondrement du coût traditionnellement associé à l'archivage des données. Autrefois, ces processus nécessitaient des efforts considérables pour archiver et transmettre ces informations aux générations suivantes. Seules les informations jugées dignes d'intérêt faisaient l'objet de cette transmission. Ainsi, comme le rappelait Viktor Mayer-Schönberger : « *Depuis le début des temps, pour les humains, l'oubli a été la norme et la mémoire l'exception. Désormais, les technologies numériques et les réseaux mondiaux ont modifié cet équilibre. Aujourd'hui, avec la massification des*

¹⁷ « Education is the point at which we decide whether we love the world enough to assume responsibility for it and by the same token save it from that ruin which, except for renewal, except for the coming of the new and young, would be inevitable. And education, too, is where we decide whether we love our children enough not to expel them from our world and leave them to their own devices, nor to strike from their hands their chance of undertaking something new, something unforeseen by us, but to prepare them in advance for the task of renewing a common world. » Hannah Arendt, *Between Past and Future* (pp. 252-253) (Viking Press 1961) et sa traduction française : *La crise de la culture* (Folio essais 1972)

*technologies, l'oubli est devenu l'exception, et la mémoire la norme*¹⁸ ». Aujourd'hui, au-delà de l'archivage des informations, l'effondrement du coût de stockage des données et du traitement en masse de ces données permet aux IA d'occuper une place sans équivalent dans le recueil, le remodelage et la transmission des connaissances humaines. Une autre tendance de fond se fait jour sur Internet ; la montée en puissance des contenus (images, textes et vidéos) générés par des IA. Leur volume est sur le point de dépasser celui des contenus créés par les humains¹⁹. Ainsi, des citoyens peuvent être progressivement influencés par des IA sans même avoir conscience que les textes ou les images consultées ont été "générés" et non pas "conçus" par des humains. Ce phénomène pourrait être d'autant plus préoccupant qu'à mesure que les IA génératives sont entraînées sur des contenus issus d'autres IA génératives, leurs résultats perdent en précision. À terme, cette tendance pourrait altérer le fonctionnement et l'efficacité des modèles d'IA génératives et certains auteurs évoquent dans ce cas la notion "*d'effondrement des modèles*" (*models collapse*)²⁰. De plus, cette "cannibalisation" des contenus peut accentuer les biais, les erreurs ou "hallucinations" générées par des IA ou encore favoriser des formes de manipulations. En effet, en plus des concepteurs des modèles d'IA génératives, qui peuvent introduire des biais dans les réponses en fonction de leurs préférences idéologiques, des acteurs malveillants peuvent créer massivement des contenus sur Internet pour saturer les données d'entraînement des modèles et ainsi modifier leurs réponses ultérieures auprès de millions d'utilisateurs.

Ainsi, avec le développement des technologies d'IA, au-delà de la mémoire, ce sont aussi les mécanismes de la transmission et la conception même de notre histoire et de notre culture qui pourraient être durablement modifiés dans nos sociétés. Ce sont également les valeurs et principes démocratiques qui pourraient

¹⁸ Since the beginning of time, for us humans, forgetting has been the norm and remembering the exception. Because of digital technology and global networks, however, this balance has shifted. Today, with the help of widespread technology, forgetting has become the exception, and remembering the default.

Viktor Mayer-Schönberger: *Delete: The Virtue of Forgetting in the Digital Age* (Princeton University Press 2009)

¹⁹ More Articles Are Now Created by AI Than Humans (Graphite.AI Nov 2024)

<https://graphite.io/five-percent/more-articles-are-now-created-by-ai-than-humans>

²⁰ AI models collapse when trained on recursively generated data.(Shumailov, I., Shumaylov, Z., Zhao, Y. et al. Nature 631, 755–759 (2024). <https://www.nature.com/articles/s41586-024-07566-y>

être remis en cause par un abandon de la souveraineté sur nos cerveaux au profit des IA qui nous « soulageraient » d'une part toujours plus grande de nos vies sociales, culturelles et politiques. Un mélange de résignation et de fascination face

“ *Si Victoria Woodcock, la responsable du logiciel [de microciblage] employé dans la campagne [du Brexit], avait été renversée par un autobus, le Royaume-Uni serait resté dans l'Union européenne...*

Dominic Cummings

Coordinateur de la campagne du Brexit

aux technologies de l'IA pourrait inciter nos sociétés à se reposer entièrement sur ces nouveaux oracles et leur confier des rôles potentiellement dangereux dans l'organisation des débats démocratiques voire dans le déroulement des processus électoraux. L'impact de ces technologies sur le débat public et plus largement sur nos démocraties reste difficile à déterminer avec précision, cependant à défaut de faire basculer d'importantes proportions du corps électoral, elles pourraient avoir des conséquences cruciales sur les scrutins en aidant à manipuler les franges d'électeurs indécis

qui, dans les pays démocratiques, déterminent le plus souvent le résultat des élections. C'est en particulier en utilisant ces technologies de microciblage des électeurs via des systèmes d'analyse de données en masse que les responsables de la campagne du *Brexit* ont pu convaincre ces électeurs indécis. Ainsi, dans son ouvrage de référence, « *Les ingénieurs du chaos* », Giuliano da Empoli évoque les méthodes utilisées par Dominic Cummings, le coordinateur de la campagne du *Brexit*, pour convaincre les électeurs, en recourant au microciblage pour "polariser" les différents segments de la population britannique :

« Pour les nouveaux docteurs Folamour de la politique, le jeu ne consiste plus à unir les gens autour du plus petit dénominateur commun mais, au contraire, à enflammer les passions du plus grand nombre possible de groupuscules pour ensuite les additionner, même à leur insu. Pour conquérir une majorité, ils ne vont pas converger vers le centre, mais joindre les extrêmes. Grâce au travail d'une équipe de scientifiques, Cummings a pu cibler des millions d'électeurs

indécis dont ses adversaires ne soupçonnaient même pas l'existence, en leur adressant exactement les messages qu'il fallait, au moment où il le fallait, pour les faire basculer dans le camp du Brexit. Mesurer l'impact précis de cette activité complexe sur le vote est impossible. Mais tout laisse penser qu'il a été important. Cummings lui-même a écrit que « Si Victoria Woodcock, la responsable du logiciel employé dans la campagne, avait été renversée par un autobus, le Royaume-Uni serait resté dans l'Union européenne ²¹ ».

C. Régulation européenne de l'IA : des objectifs industriels et politiques

La création par l'Union européenne d'un règlement sur l'intelligence artificielle (*Artificial Intelligence Act* ou *AI Act*, adopté par le Conseil européen en mai 2024) marque une rupture dans la conception même des technologies critiques dans nos sociétés. *L'AI Act* introduit une logique d'encadrement des pratiques à risque pour les citoyens de l'Union européenne. Le classement de ces risques va de risque limité pour les jeux vidéo, jusqu'à risque inacceptable pour les systèmes de notation sociale ou les IA dédiées à la manipulation politique. Il s'agit également, par une plus grande transparence, de fournir une meilleure information aux usagers sur les pratiques à risque, mais aussi d'inciter les développeurs d'IA à prendre en compte ces contraintes dans la conception de leurs services. Cela concernera particulièrement les services qui auront des répercussions sur la sécurité des personnes ou leurs libertés fondamentales. Enfin, *l'AI Act* aura pour conséquence la création d'activités économiques et donc d'emplois dans le domaine de l'éthique de l'IA, tant auprès des entreprises que des administrations.

En outre, ce règlement européen vise à éviter l'évolution vers un contrôle autoritaire des populations semblable au *Crédit social* chinois. Ce système de contrôle politique et social s'appuie sur les données recueillies sur le comportement des citoyens chinois, y compris, à terme, sur leurs données génétiques. *L'AI Act* européen rejoint ainsi les recommandations de l'*Unesco* en

²¹ da Empoli, Giuliano. *Les ingénieurs du chaos* (pp. 162-163). JC Lattès 2019

matière d'éthique de l'IA qui rappelait son attachement aux libertés et aux principes démocratiques : *« La vie privée est un droit. Cette recommandation appelle à mettre en place des garanties appropriées pour protéger ce droit, notamment en abordant des questions telles que la surveillance. Elle stipule explicitement que les systèmes d'IA ne doivent pas être utilisés à des fins de notation sociale ou de surveillance de masse²². »*

Avec l'évolution des IA, une autre liberté longtemps considérée comme un droit inaliénable pourrait être remise en cause : la liberté de tenir pour privées et secrètes les pensées des individus. Des expérimentations ont ainsi été menées pour capter à même le cerveau les émotions ou les réactions physiologiques des personnes. Ainsi, en Chine, les autorités imposent à certains travailleurs de porter des casques dotés de capteurs électroencéphalographiques pour connaître leur état émotionnel.²³

Par ailleurs, d'autres mécanismes plus discrets, mais aussi plus répandus et plus précis, permettent déjà de découvrir les convictions des personnes par l'analyse en masse des données relatives à leurs comportements. Ces techniques initialement conçues pour influencer les usagers à des fins publicitaires ont, dans un second temps, été utilisées à des fins de manipulations idéologiques politiques ou religieuses. Comme le fait remarquer l'historien Yuval Harari, dans le domaine de ce que l'on nomme désormais la « guerre cognitive », certains types de données personnelles, comme les données médicales, pourraient occuper une place déterminante.

« Il est important de se rappeler que colère, joie, ennui et amour sont aussi des phénomènes biologiques comme la fièvre ou la toux. La même technologie qui identifie la toux pourrait également identifier les rires. Si entreprises et

²² Privacy is a right. The Recommendation calls to put in place appropriate safeguards to protect this right, including addressing concerns such as surveillance. It explicitly states that AI systems should not be used for social scoring or mass surveillance purposes.

UNESCO Recommendation on the Ethics of Artificial Intelligence (23 Nov 2021)

www.ohchr.org/sites/default/files/2022-03/UNESCO.pdf

²³ With brain-scanning hats, China signals it has no interest in workers' privacy (MIT Tech Review, 30 Apr 2018) www.technologyreview.com/2018/04/30/143155/with-brain-scanning-hats-china-signals-it-has-no-interest-in-workers-privacy/

gouvernements commencent à collecter nos données biométriques en masse, ils pourraient mieux nous connaître que nous ne nous connaissons nous-mêmes, et ils pourraient non seulement prédire nos sentiments mais aussi les manipuler et nous vendre tout ce qu'ils veulent, qu'il s'agisse d'un produit ou d'un politicien. La surveillance biométrique ferait ressembler le piratage de données de Cambridge Analytica à des outils de l'âge de pierre. Imaginez la Corée du Nord en 2030, où chaque citoyen devrait porter un bracelet biométrique 24 heures sur 24. Si vous écoutez un discours du Grand Leader et que le bracelet détecte des signes révélateurs de colère, vous êtes fichu...²⁴ ».

Ainsi, à partir des informations rassemblées sur les personnes, il devient possible d'agir sur une multitude de signaux en apparence insignifiants susceptibles d'avoir des répercussions importantes sur les émotions et les processus de décision des individus. Ces signaux imperceptibles pourraient se révéler plus efficaces que des messages directs les incitant à adopter tel ou tel comportement.

L'intégration tardive des IA génératives dans *l'AI Act* montre la difficulté à élaborer des textes de régulation sur des technologies en évolution aussi rapide. Ces technologies d'IA génératives n'en étaient qu'à leurs premiers stades de développement commercial lorsque le texte de *l'AI Act* a été conçu. Plus encore, ce sont les procédures et modalités d'application de ces textes qui permettront d'en valider la portée dans les années à venir. Cependant, une forme de consensus commence à exister sur la nécessité de maîtriser, par la régulation, les risques issus de ces technologies. Les acteurs technologiques eux-mêmes reconnaissent que l'absence de régulation de l'IA conduira à des excès et des dérives sociales et politiques qui pourraient être *in fine* préjudiciables à l'industrie des technologies tout entière. Ainsi, Mustafa Suleyman, le cofondateur de *DeepMind*, reconnaît

²⁴ It is crucial to remember that anger, joy, boredom and love are biological phenomena just like fever and a cough. The same technology that identifies coughs could also identify laughs. If corporations and governments start harvesting our biometric data en masse, they can get to know us far better than we know ourselves, and they can then not just predict our feelings but also manipulate our feelings and sell us anything they want — be it a product or a politician. Biometric monitoring would make Cambridge Analytica's data hacking tactics look like something from the Stone Age. Imagine North Korea in 2030, when every citizen has to wear a biometric bracelet 24 hours a day. If you listen to a speech by the Great Leader and the bracelet picks up the tell-tale signs of anger, you are done for.

Yuval Noah Harari: the world after coronavirus (Financial Times, 20 Mar 2020)
www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75

l'importance et le caractère pionnier de la régulation européenne (*AI Act*) pour aider à maîtriser (ou endiguer) les dérives possibles de ces technologies.

« L'endiguement [de l'IA] ne constitue pas un programme qu'un gouvernement, ou même un groupe de gouvernements, peut réaliser seul. Il exige de l'inventivité et de l'audace dans le partenariat entre secteurs public et privé, et exige un ensemble de motivations tout à fait nouveau pour toutes les parties impliquées. Des législations comme l'AI Act de l'UE esquissent au moins un monde dans lequel l'endiguement trouve sa place, un monde dans lequel des gouvernements importants prennent les risques de prolifération au sérieux et témoignent d'un engagement et d'une disposition renouvelés à consentir de vrais sacrifices. La régulation ne suffit pas, mais au moins c'est un début. Prendre des mesures audacieuses, comprendre intimement les enjeux de la vague à venir : dans un monde où l'endiguement semble impossible, ce genre d'initiatives augure un avenir où il pourrait devenir concevable²⁵. »

Les régulations envisagées pour limiter les risques de dérives de l'IA (en particulier en Europe l'*AI Act*) portent essentiellement sur les applications des technologies plutôt que sur leurs fondements technologiques. Dans certains cas cependant, cette distinction devient difficile. C'est le cas par exemple de certaines applications qui permettent de créer des « *deepfakes* » ultra-réalistes. Ces technologies peuvent parfois avoir des conséquences graves sur la réputation voire sur la santé mentale ou physique des personnes. On peut citer les applications d'IA conçues pour créer des images ou des vidéos pornographiques à partir d'images issues des réseaux sociaux. Le caractère toxique de ces applications est tel que même la très influente

²⁵ Suleyman, Mustafa; Bhaskar, Michael. *La Déferlante : Technologie, pouvoir et le dilemme majeur du XXI^e siècle* (Documents) (p. 273). Fayard 2023. Ouvrage original : Suleyman, Mustafa; Bhaskar, Michael. *The Coming Wave: The instant Sunday Times bestseller from the ultimate AI insider* (p. 232). Random House 2023

The reality is that containment is not something that a government, or even a group of governments, can do alone. It requires innovation and boldness in partnering between the public and the private sectors and a completely new set of incentives for all parties. Regulations like the EU AI Act do at least hint at a world where containment is on the map, one where leading governments take the risks of proliferation seriously, demonstrating new levels of commitment and willingness to make serious sacrifices. Regulation is not enough, but at least it's a start. Bold steps. A real understanding of the stakes involved in the coming wave. In a world where containment seems like it's not possible, all of this gestures toward a future where it might be.

revue *MIT Technology Review* recommandait déjà leur interdiction en 2021²⁶. Plus récemment, la mise à disposition de ce type d'application par *Grok*, l'application d'IA d'Elon Musk, a suscité un tollé parmi les utilisateurs, ainsi qu'auprès de responsables politiques au sein de l'administration Trump²⁷. Cependant, si des limitations ou des interdictions partielles sont envisagées, une interdiction totale pourrait encore se heurter à des difficultés technologiques et juridiques²⁸.

²⁶ A horrifying new AI app swaps women into porn videos with a click (MIT Tech Review 13 Sept 2021) www.technologyreview.com/2021/09/13/1035449/ai-deepfake-app-face-swaps-women-into-porn/

²⁷ David Lammy: JD Vance agrees that sexualised AI images on X are 'unacceptable' (The Guardian 10 Jan 2026) www.theguardian.com/us-news/2026/jan/10/ai-generated-sexualised-images-x-jd-vance-grok

²⁸ Bans on deepfakes take us only so far—here's what we really need (MIT Technology Review 27 Feb 2024) www.technologyreview.com/2024/02/27/1089010/bans-on-deepfakes-take-us-only-so-far-heres-what-we-really-need/

IV. DES LIBERTÉS FONDAMENTALES EN PÉRIL ?

Dans son ouvrage sur la protection de la liberté de pensée, Susie Alegre spécialiste des droits de l'homme, revient sur la nécessité d'analyser le texte de la *Convention européenne des droits de l'homme* à la lumière du développement des nouvelles générations d'IA.

Les droits à la liberté de pensée, de conscience, de religion et de conviction, ainsi que la liberté d'opinion sont des droits absolus protégés par le droit international. Sans liberté de pensée ou d'opinion, il n'y a ni humanité ni démocratie. Pour garantir pleinement ces droits, trois conditions sont nécessaires :

- 1. la capacité de garder ses pensées privées ;*
- 2. la liberté de ne pas être manipulé ;*
- 3. que personne ne puisse être pénalisé pour ses seules pensées.²⁹*

En effet, les évolutions des technologies numériques et de l'IA peuvent apporter de nouvelles limitations à l'exercice de ces libertés et droits fondamentaux. Désormais, il est possible de « deviner » les pensées ou les convictions d'une

²⁹ "The rights to freedom of thought, conscience, religion and belief and freedom of opinion are absolute rights protected in international law. Without freedom of thought or opinion, we have no humanity, and we have no democracy. Making these rights real requires three things:

1. the ability to keep your thoughts private;
2. freedom from manipulation of your thoughts;
3. that no one can be penalised for their thoughts alone."

See Vermeulen, B., 'Article 9', in P. van Dijk et al. (eds), *Theory and Practice of the European Convention on Human Rights*, 4th edn, Cambridge, Intersentia Press, 2006, p.752.

Susie Alegre. *Freedom to Think: Protecting a Fundamental Human Right in the Digital Age* (Atlantic Books 2023)

personne à partir d'actions apparemment anodines (par exemple sur les réseaux sociaux). Comme l'ont démontré des chercheurs de l'université de *Cambridge* en 2013, l'analyse des « *likes* » sur *Facebook* permettait déjà d'identifier avec une grande précision des données sensibles sur les individus³⁰. Les algorithmes utilisés dans cette étude se sont révélés précis à 88 % pour déterminer l'orientation sexuelle des personnes, à 95 % pour distinguer les Afro-Américains des personnes d'origine caucasienne, et à 85 % pour différencier les convictions politiques entre républicains et démocrates. Les croyances religieuses (chrétiennes ou musulmanes) ont été correctement identifiées dans 82 % des cas et le statut relationnel des utilisateurs et la toxicomanie ont été prédits avec une précision comprise entre 65 % et 73 %. Une étude dirigée en 2014 par Michal Kosinski à l'université *Stanford* démontrait même que les profils psychologiques établis par des modèles d'IA basés sur les données issues des réseaux sociaux étaient plus précis que ceux établis par des proches (amis, famille, conjoint ou collègues)³¹.

Quant à la liberté de ne pas être manipulé, elle aussi est désormais remise en question avec la « démocratisation » de certains modèles d'IA. Ces modèles permettent d'analyser les caractéristiques des personnes à partir des nombreuses informations issues des réseaux sociaux puis de les faire réagir et les orienter grâce à des messages ciblés. La première affaire célèbre concernant l'utilisation frauduleuse des données issues de *Facebook* à des fins de manipulations politiques est liée à la société *Cambridge Analytica*³². Cette affaire, dans laquelle l'IA a joué un rôle central³³, pourrait avoir joué un rôle crucial dans l'élection présidentielle américaine de 2016. Cette élection s'est en effet jouée à 107 000 voix près dans

³⁰ Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. *The Proceedings of the National Academy of Sciences (PNAS)* Apr 9 2013
www.pnas.org/doi/full/10.1073/pnas.1218772110

³¹ Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. *The Proceedings of the National Academy of Sciences (PNAS)* 2015 Jan 27
www.pnas.org/doi/10.1073/pnas.1418680112

³² Cambridge Analytica: how did it turn clicks into votes? (*The Guardian*, 6 May 2018)
www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie

³³ Decoded: How Cambridge Analytica used AI (*Politico* 28 Jan 2020)
www.politico.eu/newsletter/ai-decoded/politico-ai-decoded-how-cambridge-analytica-used-ai-no-google-didnt-call-for-a-ban-on-face-recognition-restricting-ai-exports/

3 États (Pennsylvanie, Wisconsin, Michigan), ce qui ne représentait alors que 0,09 % des suffrages exprimés³⁴.

A. IA & désinformation : les nouvelles menaces sur la démocratie

La liberté d'information est celle qui permet de vérifier l'existence de toutes les autres.

Christophe Deloire
Directeur général de *Reporters sans frontières*³⁵

Les évolutions des technologies d'IA associées aux réseaux sociaux ont déjà des effets considérables sur l'ensemble des éléments qui constituent nos démocraties. Cela va de la remise en cause des informations nécessaires à l'exercice du libre arbitre des citoyens par la création de « *deepfakes* », à la coordination sur les réseaux sociaux de mouvements d'opinion artificiels, en passant par l'envoi de messages ciblés en fonction des profils précis établis par les courtiers en données. Chacune des composantes de nos démocraties peut désormais être attaquée par des acteurs politiques, des États ou des groupes hostiles.

La nouveauté des IA réside dans leur capacité d'intrusion dans ce qui semblait inaccessible par les moyens traditionnels de la propagande : le fonctionnement même de la conscience humaine. Ainsi, pour Simon McCarthy-Jones, professeur de psychologie au *Trinity College* de Dublin, l'altération de cette part de l'esprit

³⁴ How Trump won the presidency with razor-thin margins in swing states (Washington Post 11 Nov 2016) www.washingtonpost.com/graphics/politics/2016-election/swing-state-margins/

³⁵ Christophe Deloire "La liberté d'information est celle qui permet de vérifier l'existence de toutes les autres" (INA 2012) mediaclip.ina.fr/fr/i22075282-christophe-deloire-la-liberte-d-information-est-celle-qui-permet-de-verifier-l-existence-de-toutes-les-autres.html

humain qui participe à l'élaboration de ses convictions équivaldrait à le priver de ce qui le rend apte à participer aux processus démocratiques.

Perdre la souveraineté sur nos esprits, c'est perdre notre dignité, notre démocratie, et même nous perdre nous-mêmes. Cette souveraineté est souvent désignée par l'expression d'autonomie mentale. Il s'agit de « la capacité spécifique de contrôler ses propres fonctions mentales », incluant des aspects tels que l'attention, la mémoire, la planification, la pensée rationnelle et la prise de décision. La dignité, « la présomption selon laquelle qu'un individu est une personne dont les actions, les pensées et les préoccupations méritent un respect inhérent au fait qu'elles ont été choisies, organisées et guidées », suppose une autonomie mentale. Les démocraties, où les citoyens choisissent les lois qui les contraignent, ne peuvent exister que si les citoyens sont mentalement autonomes. La capacité de penser librement est si essentielle à notre identité que la violer revient à nous priver « de notre personnalité tout entière »³⁶.

Sur les réseaux sociaux, les internautes sont ciblés par des algorithmes qui s'efforcent de proposer des contenus correspondant à leurs goûts, centres d'intérêt, préférences partisans, etc. Il en résulte une tendance à l'enfermement dans des univers clos (bulles informationnelles) et un risque désormais avéré de manipulations politiques, avec des messages jouant principalement sur l'émotion et la démagogie. Ce fut notamment le cas aux États-Unis avec l'affaire *Cambridge Analytica*, ou lors de l'assaut du *Capitole*, où des groupes extrémistes comme *QAnon* se sont structurés grâce aux algorithmes de « microciblage » de *Facebook*. La spécialiste du droit international Anu Bradford a ainsi commenté les

³⁶ To lose sovereignty over our minds is to lose our dignity, our democracy, and even our very selves. Such sovereignty is termed mental autonomy. This is “the specific ability to control one's own mental functions,” which include attention, memory, planning, rational thought and decision making (Metzinger, 2013). Dignity, “the presumption that one is a person whose actions, thoughts and concerns are worthy of intrinsic respect, because they have been chosen, organized and guided” (Nuffield Council on Bioethics, 2002, p. 121) requires mental autonomy. Democracies, in which citizens choose the laws that bind them (Johnson and Cureton, 2019), are only possible if citizens are mentally autonomous. The ability to think freely is so essential to our identity that to violate it is to deprive us “of personhood altogether” (Halliburton, 2009, p. 868).

McCarthy-Jones S. The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century (Frontiers in Artificial Intelligence. 26 Sep 2019) www.ncbi.nlm.nih.gov/pmc/articles/PMC7861318/

conséquences de cet assaut sur la politique américaine et sur la régulation des plateformes :

Selon certains commentateurs, l'attaque du 6 janvier 2021 contre le Capitole des États-Unis pourrait avoir marqué un tournant dans l'approche américaine de la réglementation des contenus en ligne. Emily Bazelon a décrit l'événement comme une démonstration de l'échec du « marché américain des idées », ce qui implique que les croyances techno-libertaires américaines sur la liberté d'expression ont été parmi les causes de l'événement. Les émeutes du Capitole ont conduit à une large condamnation du rôle joué par les plateformes en ligne, qui ont permis à des discours nuisibles et dangereux de prendre une ampleur telle qu'ils ont finalement abouti à la violence.³⁷

Plus récemment, les émeutes survenues après les meurtres d'enfants à Southport en Angleterre ont aussi donné lieu à une réflexion sur le rôle des médias sociaux dans la « polarisation » *post-Brexit* de l'opinion publique britannique³⁸. Au-delà de la désinformation et des campagnes de messages ciblés, d'autres pratiques de l'IA pourraient subvertir le fonctionnement même de nos institutions démocratiques. La combinaison des données issues des réseaux sociaux et la puissance de traitement des systèmes d'IA peuvent modifier subtilement et de manière indétectable les conditions d'exercice du vote des citoyens. Pour Elena Kagan, juge à la *Cour suprême* américaine, les redécoupages électoraux (*gerrymandering*) assistés par intelligence artificielle pourraient saper les fondements même de la démocratie. Ces manipulations permettraient de rendre certaines circonscriptions électorales « imperdables » et ce de manière imperceptible pour les citoyens. Elle

³⁷ Some commentators suggest that the January 6, 2021, attack on the US Capitol may have been a turning point in the US's approach toward regulating online content.¹⁰ Emily Bazelon described the event as showing how the “American marketplace of ideas clearly failed,” implying that American techno-libertarian beliefs about free speech were among the causes of the event. The Capitol riots led to broad condemnation of the role that online platforms played in allowing harmful and dangerous speech to gather so much momentum that it ultimately led to violence.

Anu Bradford. *Digital Empires_The Global Battle to Regulate Technology* (p. 363) (Oxford University Press 2023)

³⁸ ‘A polarisation engine’: how social media has created a ‘perfect storm’ for UK’s far-right riots (Carole Cadwalladr - The Guardian 3 Aug 2024) www.theguardian.com/media/article/2024/aug/03/a-polarisation-engine-how-social-media-has-created-a-perfect-storm-for-uks-far-right-riots

déclare : « À mesure que le temps passe, les manipulations liées aux redécoupages électoraux ne feront que s'aggraver. Ce qui était faisable avec un papier et un stylo deviendra insignifiant face à ce qui deviendra possible avec le développement de l'intelligence artificielle. À mesure que les données deviendront plus précises, et que s'amélioreront les techniques d'analyse, quelque part en chemin, la souveraineté du peuple aura disparu...³⁹ ».

“ Pour un coût dérisoire, il devient possible de créer des robots de désinformation autonomes qui utilisent des systèmes d'intelligence artificielle pour répondre à des utilisateurs en adaptant leurs réponses à des millions de profils différents...”

Récemment, les technologies des intelligences artificielles génératives ont permis de créer et de personnaliser à moindre coût les messages envoyés à de vastes groupes d'utilisateurs des services de l'Internet. Ces nouvelles capacités ont permis la « démocratisation » de l'ingérence et de la manipulation automatisées de l'opinion publique. En plus des entités étatiques, ces technologies de désinformation sont ainsi devenues accessibles aux acteurs privés, à des

groupes politiques ou même des individus isolés. Pour un coût dérisoire, il devient possible de créer des « robots de désinformation autonomes »⁴⁰. En utilisant des systèmes d'IA, ces robots peuvent répondre spécifiquement à des millions d'utilisateurs en adaptant les réponses à leurs profils. Les personnes visées par ces campagnes croient être confrontées à des personnes réelles qui partagent les mêmes opinions. Plus récemment, plusieurs études scientifiques ont démontré que les messages issus des grands modèles de langage (*LLM*) ont un plus grand impact

³⁹ Supreme Court Justice Elena Kagan warns AI-powered gerrymandering could undermine US democracy (Business Insider, 28 Jun 2019)
www.businessinsider.com/justice-elena-kagan-warns-ai-powered-gerrymandering-may-hurt-democracy-2019-6

⁴⁰ Inside Countercloud : a Fully Autonomous AI Disinformation System (The Debrief, 16 août 2023)
<https://thedebrief.org/countercloud-ai-disinformation/>

sur les convictions politiques de leurs usagers que les méthodes traditionnelles et en particulier les publicités politiques⁴¹. Cette inquiétude pour l'avenir de la démocratie est également partagée par Ben Buchanan, ancien conseiller pour l'IA à la *Maison-Blanche* :

De manière encore plus insidieuse, l'IA offre la possibilité d'automatiser les campagnes de propagande et de désinformation, ajoutant ainsi de l'huile sur le feu. L'IA peut écrire des textes qui semblent authentiques, générer des vidéos qui semblent réelles, et le faire plus rapidement et à moindre coût que n'importe quel humain ne pourrait le faire ou le détecter. L'IA façonne également le terrain sur lequel se déroulent les campagnes de désinformation. Elle détermine les histoires qui apparaissent sur les fils d'actualité de Facebook, les tweets qui apparaissent sur les chronologies X (ex-Twitter) et les vidéos qui apparaissent ensuite dans la lecture automatique de YouTube. Les grandes plateformes Internet sont devenues des arènes où des visions du monde s'affrontent chaque jour, avec des acteurs malveillants qui tentent d'appriivoiser les algorithmes des plateformes pour que leurs messages deviennent viraux, et des entreprises qui tentent – et échouent souvent – d'utiliser l'IA pour tenir à distance la désinformation et les autres formes de haine⁴².

⁴¹ Chatbots Can Meaningfully Shift Political Opinions, Studies Find (New York Times, 21 Dec 2025) www.nytimes.com/2025/12/05/well/ai-chatbots-political-opinions-study.html

⁴² Even more insidiously, AI offers the prospect of automating propaganda and disinformation campaigns, adding more fuel to the new fire. AI can write text that seems genuine, generate videos that seem real, and do it all faster and cheaper than any human could match or detect. AI also shapes the terrain on which disinformation efforts unfold. It influences which stories appear on Facebook news feeds, which tweets show up on Twitter timelines, and which video pops up next on YouTube's autoplay. Major internet platforms have become arenas where worldviews clash every day, with malicious actors trying to coax corporate algorithms to make their messages go viral, and companies trying—and often failing—to use AI to keep disinformation and other forms of hate at bay.

Buchanan, Ben; Imbrie, Andrew. *The New Fire: War, Peace, and Democracy in the Age of AI* (p. 8) (MIT Press 2022)

B. Des formes inédites de manipulations politiques

Ces manipulations politiques pourraient aussi être encore amplifiées dans les nouveaux univers virtuels partagés qu'imaginent les promoteurs du *Métavers*. Les expériences auxquelles participeront leurs utilisateurs reposeront massivement sur des IA qui permettront de concevoir à grande échelle des contenus et des services hyper-individualisés. Cette nouvelle génération de services qui associeront réalité virtuelle et contenus issus d'IA génératives pourrait aider à disséminer de nouvelles formes de désinformation de masse. Ainsi dans un article du *Washington Post* au titre évocateur ; « *La désinformation sur Facebook est déjà assez grave. Le métavers sera encore pire* », le chercheur en IA, Rand Waltzman, envisage ce à quoi les campagnes politiques « améliorées par l'IA » pourraient ressembler dans ces nouveaux univers :

Un candidat politique fait un discours devant des millions de personnes. Alors que chaque spectateur pense voir la même version du candidat, dans la réalité virtuelle, chacun voit en réalité une version légèrement différente. Le visage du candidat a été subtilement modifié pour ressembler à chaque spectateur en mêlant les traits du visage de chaque spectateur avec ceux du candidat. Les spectateurs ne sont pas conscients d'une quelconque manipulation de l'image. Pourtant, ils sont fortement influencés par elle : chaque personne du public est plus favorablement disposée envers le candidat qu'il ne l'aurait été sans aucune manipulation numérique. Ce phénomène n'est pas hypothétique. On sait depuis longtemps que le mimétisme peut être exploité comme un puissant outil d'influence. Une série d'expériences menées par des chercheurs de Stanford a montré que le fait de modifier légèrement les traits d'une personnalité politique inconnue pour ressembler à chaque électeur permettait aux gens d'évaluer les politiciens plus favorablement. Les expériences ont pris des photos des participants à l'étude et de vrais candidats dans une maquette d'une campagne électorale. Les photos de chaque candidat ont été modifiées pour ressembler à chaque participant. Les études ont révélé que même si 40 % des traits du participant étaient intégrés au visage du candidat, les participants n'avaient aucune conscience de la manipulation. Dans le métavers, il est facile d'imaginer ce type de mimétisme à grande échelle.⁴³

⁴³ A political candidate is giving a speech to millions of people. While each viewer thinks they are seeing the same version of the candidate, in virtual reality they are actually each seeing a slightly different version. For each and every viewer, the candidate's face has been subtly modified to resemble the viewer. This is done by blending

Cette représentation des candidats « améliorés par l'IA » pourrait n'être que la première étape vers l'utilisation de personnages entièrement « synthétiques » conçus via des IA génératives. Il deviendrait alors possible de convaincre les citoyens sans que ceux-ci n'aient conscience d'être confrontés à des personnages entièrement artificiels. Ce scénario de science-fiction ressemblerait au film *Simone* d'Andrew Niccol en 2002 dans lequel un réalisateur désargenté, joué par Al Pacino, créait une star de cinéma entièrement artificielle. Les interactions quotidiennes avec des personnages artificiels pourraient ainsi devenir réalité. Avec les dernières générations d'IA génératives, ces compagnons artificiels commencent déjà à devenir des réalités commerciales. Certains services permettent de créer des partenaires virtuels, sous forme d'avatars photoréalistes, qui s'appuient sur des IA génératives pour leur apparence et aussi pour générer des interactions avec leurs utilisateurs.⁴⁴ La frontière entre humain et machine, que certains acteurs technologiques envisagent d'abolir, pourrait devenir de plus en plus difficile à discerner. Comme le décrit Laurence Devillers, professeure en informatique appliquée aux sciences sociales à l'Université Paris-Sorbonne : « *Nous entrons dans une ère de relations inextricables entre l'humain et les machines, une relation de confiance et d'affection au sein de laquelle la séparation entre vivant et artefact, aujourd'hui si nette, deviendra de plus en plus floue*⁴⁵. »

features of each viewer's face into the candidate's face. The viewers are unaware of any manipulation of the image. Yet they are strongly influenced by it: Each member of the audience is more favorably disposed to the candidate than they would have been without any digital manipulation. This is not speculation. It has long been known that mimicry can be exploited as a powerful tool for influence. A series of experiments by Stanford researchers has shown that slightly changing the features of an unfamiliar political figure to resemble each voter made people rate politicians more favorably. The experiments took pictures of study participants and real candidates in a mock-up of an election campaign. The pictures of each candidate were modified to resemble each participant. The studies found that even if 40 percent of the participant's features were blended into the candidate's face, the participants were entirely unaware the image had been manipulated. In the metaverse, it's easy to imagine this type of mimicry at a massive scale.

Rand Waltzman. Facebook misinformation is bad enough. The metaverse will be worse (Washington Post, 22 Aug 2022) www.washingtonpost.com/opinions/2022/08/22/metaverse-political-misinformation-virtual-reality/

⁴⁴ The Perfect Girlfriend (Esquire, 3 Oct 2024) www.esquire.com/news-politics/a62452522/ai-girlfriend/

⁴⁵ Laurence Devillers. Les robots émotionnels: Santé, surveillance, sexualité... : et l'éthique dans tout ça ? (p.40). Humensis 2019 voir également ses travaux sur la détection des émotions par les IA :

C. Vers des relations affectives avec les IA ?

Un exemple récent de création d'IA générative « séductrice » s'est produit à l'occasion du lancement par la société *OpenAI* de la version vocale de son assistant *Chat GPT-4o*. La voix de cet assistant dénommée « *Sky* » semblait directement inspirée de celle du film « *HER* » de Spike Jonze. Dans ce film de 2013, le personnage joué par Joaquin Phoenix tombait amoureux de la voix féminine du système d'exploitation de son ordinateur interprétée par Scarlett Johansson. La voix « *Sky* » de l'assistant *GPT-4o* ressemblait tellement à la sienne que l'actrice a souhaité lancer une action en justice contre *OpenAI* pour atteinte à son image.⁴⁶ Ce qui a par la suite conduit la société *OpenAI* à retirer la voix *Sky* de son assistant virtuel. Dans son communiqué Scarlett Johansson, se plaignait de l'attitude cavalière de Sam Altman, le PDG de la société *OpenAI*, qui après avoir essuyé plusieurs refus de la part de l'actrice avait quand même décidé d'utiliser la voix *Sky*. Dans son communiqué, l'actrice se prononçait également pour que des solutions juridiques soient élaborées en particulier pour protéger les métiers artistiques : « *À une époque où nous sommes tous confrontés à des deepfakes et à des questions liées à la protection de notre propre image, de notre propre travail, de notre propre identité, je crois que ces questions méritent une clarté absolue. J'attends avec impatience une solution qui nécessitera de la transparence et l'adoption d'une législation appropriée pour contribuer à garantir la protection des droits individuels*⁴⁷. »

En septembre 2024, du seul fait dans ce conflit avec *OpenAI*, Scarlett Johansson a intégré la liste des « *100 personnes les plus influentes de l'IA* » du magazine *Time*.⁴⁸

Laurence Devillers, Laurence Vidrascu, Lori Lamel. Challenges in real-life emotion annotation and machine learning based detection, *Neural Networks*, Volume 18, Issue 4, 2005, Pages 407-422, www.sciencedirect.com/science/article/abs/pii/S08933608005000407?via%3Dihub

⁴⁶ Scarlett Johansson said she was forced to hire legal counsel to deal with Sam Altman and OpenAI (Fortune, 21 May 2024) fortune.com/2024/05/20/scarlett-johansson-chatgpt-sky-voice-lawyer-sam-altman-openai-her/

⁴⁷ " In a time when we are all grappling with deepfakes and the protection of our own likeness, our own work, our own identities, I believe these are questions that deserve absolute clarity. I look forward to resolution in the form of transparency and the passage of appropriate legislation to help ensure that individual rights are protected. " Scarlett Johansson's Statement About Her Interactions With Sam Altman (New York Times, 20 May 2024) www.nytimes.com/2024/05/20/technology/scarlett-johansson-openai-statement.html

⁴⁸ The 100 Most Influential People in AI 2024 - Time (5 Sep 2024)

Sur le plan technologique et industriel, cette controverse avec *OpenAI* démontre aussi à quel point les acteurs des technologies souhaitent que les utilisateurs puissent créer des relations affectives avec ces IA. Pour que les technologies d'IA puissent se développer dans la vie quotidienne de leurs usagers elles devront apparaître comme des éléments rassurants et progressivement faire partie de leur environnement. Comme le précisait le pionnier de l'informatique ubiquitaire, Mark Weiser : « *Les technologies les plus profondes sont celles qui disparaissent. Elles se fondent dans la trame de la vie quotidienne jusqu'à en devenir indiscernables* »⁴⁹. Pour les entreprises qui développeront ces IA, il est important que les utilisateurs se sentent en confiance afin qu'ils partagent davantage d'informations qui permettront à leur tour de perfectionner ces IA. Un champ entier de recherche en informatique est ainsi consacré à l'humanisation des relations homme/machine mais aussi aux risques de déshumanisation que cette relation affective avec les machines pourrait avoir en retour pour les relations avec les humains réels⁵⁰. L'une des conséquences de cette déshumanisation étant la remise en cause de la confiance entre individus et, avec elle, la remise en question de la « *conversation* » entre citoyens et de la confrontation d'opinions. Cette conversation qui, en plus de former la trame de nos vies quotidiennes, constitue l'ossature même de la démocratie.

<https://time.com/collections/time100-ai-2024/>

⁴⁹ Mark Weiser. *The Computer for the 21st Century* (Scientific American 1991)

<https://ics.uci.edu/~corps/phaseii/Weiser-Computer21stCentury-SciAm.pdf>

⁵⁰ Voir également sur ce point les travaux de Sherry Turkle au MIT sur les risques que cette "humanisation" des relations homme/machine pourraient poser aux relations interpersonnelles.

Alone Together: Why We Expect More from Technology and Less from Each Other (Basic Books 2011)

V. LES CONSÉQUENCES DES IA SUR LES DÉMOCRATIES

A. Des IA protectrices contre des IA hostiles ?

Dans les sociétés démocratiques, la détection des ingérences et la lutte contre les atteintes à la démocratie relèveront de plus en plus souvent des technologies d'intelligence artificielle. Les experts de ces nouvelles technologies d'IA pourraient ainsi être les seuls capables de vérifier si les opinions publiques ont été manipulées ou si l'expression des suffrages est restée sincère⁵¹. À terme, des pressions considérables pourraient s'exercer sur ces experts et leur rôle dans les processus démocratiques ne manquerait pas de générer des suspicions. Le contrôle par les citoyens des usages de ces technologies deviendra encore plus crucial que par le passé. Dans le même temps, il conviendra de redéfinir le périmètre des données sensibles. En effet, les évolutions des technologies d'IA permettent d'analyser des informations en apparence anodines pour déduire des données sensibles sur les individus.

Enfin, les citoyens devront aussi être mieux informés des risques des nouvelles formes de surveillance et de manipulation qui, comme on a pu le constater lors de la guerre en Ukraine, font désormais partie intégrante des nouvelles formes de belligérance que l'on nomme désormais les « guerres hybrides ». Henry Kissinger les décrit ainsi : *« Aujourd'hui, après la guerre froide, les grandes puissances et d'autres États ont renforcé leur arsenal avec des capacités cybernétiques dont l'utilité découle en grande partie de leur opacité et de leur capacité à être niée et, dans certains cas, de leur fonctionnement, à la frontière ambiguë de la*

⁵¹ Richard Ghevontian. La notion de sincérité du scrutin-Cahiers du Conseil Constitutionnel n° 13 jan 2003 www.conseil-constitutionnel.fr/nouveaux-cahiers-du-conseil-constitutionnel/la-notion-de-sincerite-du-scrutin

*désinformation, de la collecte de renseignements, du sabotage et du conflit traditionnel – créant des stratégies sans doctrines reconnues*⁵². »

Pour de nombreuses campagnes de désinformation, l'IA joue désormais un rôle clé. Qu'il s'agisse de l'analyse des profils des cibles de ces campagnes, de l'élaboration des messages, d'images ou de vidéos via des IA génératives. Des milliers voire des millions de faux comptes peuvent aussi être coordonnés pour constituer un effet de masse (que l'on nomme « *astroturfing* »). Les électeurs peuvent ainsi être influencés par ce qu'ils pensent être des mouvements spontanés de l'opinion publique. Certaines expérimentations visent ainsi à utiliser des IA génératives pour remettre en question des croyances issues de campagnes de désinformation auprès des certains électeurs⁵³. D'autres initiatives de lutte contre la désinformation et les ingérences étrangères s'inscrivent dans le cadre de la *Stratégie européenne pour les données*⁵⁴. Cette stratégie propose d'allier régulation et développement des innovations dans le domaine de l'IA avec la création *d'espaces de données (ou data spaces)*. Il s'agit pour les pays de l'UE de développer des infrastructures souveraines pour le partage de données qui permettront de développer des IA dans domaines comme l'agriculture, l'éducation, la culture, les transports, les services publics ou encore l'énergie ou l'environnement. Ces données stratégiques et en particulier les données protégées des administrations devront ainsi être sécurisées sur le plan technologique mais aussi juridique (afin de protéger ces données contre d'éventuelles cyberattaques ou des possibilités d'interception via des lois extraterritoriales).

⁵² Today, after the Cold War, the major powers and other states have augmented their arsenal with cyber capabilities whose utility derives largely from their opacity and deniability and, in some cases, their operation, at the ambiguous border of disinformation, intelligence collection, sabotage, and traditional conflict — creating strategies without acknowledged doctrines.

Henry A. Kissinger, Eric Schmidt and Daniel Huttenlocher. « The Age of AI: And Our Human Future. » (John Murray Press, 2021)

⁵³ This Chatbot Pulls People Away From Conspiracy Theories (New York Times, 12 Sep 2024)
www.nytimes.com/2024/09/12/health/chatbot-debunk-conspiracy-theories.html

⁵⁴ A European strategy for data (Communication from the Commission to the European Parliament, the Council, The European Economic and Social Committee and the Committee of the Regions, Feb 2020)
<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066>

Un exemple d'utilisation des données publiques à des fins de lutte contre la désinformation correspond au projet *Gen4Dem*⁵⁵ porté par *l'AFP* et *Visions* avec le soutien de *France Télévision* et *Reporters Sans Frontière*. Cette initiative vise à utiliser des IA génératives afin de favoriser la création et la diffusion d'information de confiance. Il s'appuie sur les *espaces de données (Data Spaces)* pour connecter des informations scientifiques, éducatives, politiques et judiciaires afin de permettre la recherche d'information et faciliter la comparaison entre les informations et ainsi être en mesure de répondre plus rapidement aux campagnes de désinformation.

Les grandes plateformes de l'Internet utilisent également des IA pour la modération des contenus et pour détecter les « comportements inauthentiques ». La principale caractéristique des réseaux sociaux est de rendre l'inscription la plus simple possible. Cette particularité permet à des systèmes automatisés d'effectuer des inscriptions en masse pour mettre en place des campagnes politiques de désinformation à grande échelle. Grâce aux IA, il devient possible de créer des millions de faux comptes en quelques jours. La lutte contre ces pratiques implique l'analyse de la dynamique des échanges entre certains comptes, sur la nature des messages échangés ainsi que sur le type de coordination de ces comptes pour amplifier certains messages. Ainsi, *Facebook* utilise des systèmes d'IA pour détecter et effacer massivement ces faux comptes. En 2023 seulement, Facebook a supprimé plus de 3 milliards de faux comptes⁵⁶.

Plutôt que de tenter de limiter « à la source » les campagnes de désinformation, certains spécialistes préconisent de s'appuyer sur des IA pour combattre leurs manifestations⁵⁷. Si ces processus de protection des débats

⁵⁵ Intelligence artificielle générative : « Ne pas agir, c'est risquer de laisser les grandes plateformes étrangères façonner notre paysage informationnel » (tribune publiée dans *Le Monde* du 13 janvier 2025) www.lemonde.fr/idees/article/2025/01/13/intelligence-artificielle-generative-ne-pas-agir-c-est-risquer-de-laisser-les-grandes-plateformes-etrangeres-faconner-notre-paysage-informationnel_6495580_3232.html

⁵⁶ Inauthentic Behavior Meta Transparency Center (2024)

<https://transparency.meta.com/fr-fr/policies/community-standards/inauthentic-behavior/>

⁵⁷ Stopping AI disinformation: Protecting truth in the digital world (World Economic Forum, 14 Jun 2024) www.weforum.org/agenda/2024/06/ai-combat-online-misinformation-disinformation/

démocratiques par l'IA venaient à se généraliser, ils pourraient eux-mêmes faire l'objet de critiques. En raison du pouvoir considérable qui serait accordé à ces IA sur le déroulement des processus électoraux, des soupçons pèseraient inéluctablement sur les concepteurs de ces technologies, ce qui conduirait à alimenter la défiance déjà importante vis-à-vis des processus électoraux et plus largement des institutions démocratiques. En un sens, cette opacité des institutions démocratiques qui seraient protégées par des IA constituerait la manifestation ultime de l'effet de « *société boîte noire* » que Frank Pasquale dénonçait dans l'ouvrage du même nom⁵⁸. Il y plaidait pour une plus grande transparence des algorithmes des grandes plateformes sous la forme d'une ouverture du code à des experts. Il évoquait dans le cas la notion de transparence qualifiée qui permettrait à des groupes d'experts indépendants d'auditer le code des algorithmes les plus importants pour le fonctionnement démocratique de nos sociétés.

B. Les nouvelles formes de lutte contre la désinformation

Parmi les mesures qui permettent de limiter l'impact des campagnes de désinformation, certains experts envisagent désormais d'agir sur la réceptivité des citoyens plutôt que de combattre les effets de ces campagnes lorsqu'elles ont eu lieu. Ainsi, le récent rapport des *États généraux de l'information* propose de développer le « *pre-bunking*⁵⁹ » qui à l'inverse du « *debunking* » vise à sensibiliser

⁵⁸ Traduction française "Black Box Society: Les algorithmes secrets qui contrôlent l'économie et l'information" par Frank Pasquale (FYP Editions 2015) ouvrage original: "Black Box Society - The Secret Algorithms That Control Money and Information" by Frank Pasquale (Harvard University Press 2015)

⁵⁹ Voir sur ce point l'ouvrage de David Colon : La guerre de l'information: Les États à la conquête de nos esprits (Taillandier 2023) ainsi que le rapport de l'OCDE "Facts not Fakes: Tackling Disinformation, Strengthening Information Integrity" (OECD, 4 Mar 2024)
www.oecd.org/en/publications/facts-not-fakes-tackling-disinformation-strengthening-information-integrity_d909ff7a-en/full-report.html

en amont les possibles cibles de la désinformation plutôt que d'intervenir en aval, lorsque ces campagnes auront déjà cheminé dans les esprits des citoyens :

Le groupe de travail considère la mise en place de techniques de pre-bunking (réfutation par anticipation) comme nécessaire pour permettre d'immuniser les citoyens aux campagnes de désinformation. L'éducation au phénomène de « weaponisation », c'est-à-dire d'instrumentalisation de l'environnement informationnel à des fins stratégiques et à des fins d'influence, est clé, car ils peuvent être des cibles potentielles pour la diffusion de mauvaises informations ou des informations biaisées. Le groupe conseille des phases de test & learn, afin de déterminer les meilleurs moyens d'immuniser les citoyens face à la désinformation. Un appui sur la recherche universitaire est capital pour déterminer les mesures les plus efficaces. À ses fonctions de détection et caractérisation des menaces, Viginum devrait se voir confier une mission de résilience, de manière à structurer les efforts de réfutation par anticipation en France et sur tous les domaines d'activité, s'inspirant du Swedish Psychological Defence Agency (mpf.se).⁶⁰

En tant qu'individus, mais aussi en tant que société, nous devons nous interroger sur le pouvoir politique et social donné à ces entreprises de modifier les formes mêmes de nos sociétés. L'IA pourrait en effet accentuer encore ce pouvoir dans les années à venir. Les modèles d'IA peuvent progressivement prendre en compte les actions des utilisateurs pour perfectionner leurs réponses. Il devient alors possible aujourd'hui d'utiliser les données massivement recueillies sur les individus pour améliorer non pas la connaissance de leurs comportements mais plutôt la manière de les façonner au cours du temps. Ainsi, qu'il s'agisse des comportements de consommation, des convictions idéologiques, religieuses ou politiques il devient possible d'intégrer plusieurs milliers de paramètres dans les messages adressés aux personnes afin de les influencer.

L'analyse du comportement des individus, via l'extraction des données personnelles, correspondait à une démarche « artisanale », désormais avec l'IA

⁶⁰ Extrait du rapport des états généraux de l'information - Protéger et développer, le droit à l'information : une urgence démocratique (12 sept 2024) www.vie-publique.fr/files/rapport/pdf/295405.pdf

c'est l'industrialisation (ou la démocratisation dans le sens le plus négatif du terme) de la manipulation des émotions de populations tout entières qui devient accessible à moindre coût. La juriste britannique Susie Alegre décrit en ces termes les risques liés à l'accumulation des données personnelles : « *Jusqu'à présent, la plupart des discussions sur les solutions se sont concentrées sur la protection de la vie privée et des données. Or, le problème fondamental de techniques comme le microciblage comportemental et le modèle [économique] du « capitalisme de surveillance » ne réside pas dans les données, mais dans la manière dont elles sont utilisées pour accéder à notre esprit*⁶¹ ».

C. Une course au moins-disant démocratique ?

L'objectif de nombreux régimes autoritaires est désormais que la démocratie soit perçue comme un simple « narratif » de surcroît moins efficace que celui des régimes non démocratiques. Parmi les arguments qu'ils opposent aux régimes démocratiques figurent l'instabilité et l'incapacité à concevoir des politiques publiques sur la longue durée. Les réseaux sociaux puis les IA sont ainsi devenus des leviers pour convaincre leur population mais aussi celle des pays démocratiques des bienfaits des régimes autoritaires. Désormais, pour ces régimes, les IA représentent des nombreux avantages dans la conduite des guerres hybrides menées contre les démocraties. Ainsi, les IA permettent de propager et d'amplifier à moindre coût des narratifs sur l'efficacité supposée des formes autoritaires tout en discréditant les régimes démocratiques et en accentuant la polarisation de leurs opinions publiques.

⁶¹ However, most of the discussion so far about the solutions has focused on privacy and data protection. But the fundamental problem with techniques like behavioural microtargeting and the 'surveillance capitalism' model is not the data; it's how it is used as a key to our minds.

Susie Alegre. Freedom to Think: Protecting a Fundamental Human Right in the Digital Age (Atlantic Books 2023)

Cependant, l'une des faiblesses des régimes autoritaires, comme l'a montré la Chine au début de la pandémie de Covid, est liée à la crainte pour les responsables locaux de transmettre des informations perçues comme une remise en cause du

“ ***L'IA améliorera les outils de l'arsenal de l'autocrate, tels les armes autonomes létales, les cyberopérations et la désinformation. Dans les autocraties, les préoccupations éthiques ne seront que des obstacles mineurs...*** ”

Ben Buchanan

pouvoir central. C'est la crainte d'être sanctionnés qui a conduit ces responsables à cacher la réalité de la pandémie pendant les premières semaines qui étaient cruciales pour la combattre.

L'IA, apparaît aussi pour les régimes autoritaires comme un début de solution à la difficulté de disposer d'informations fiables pour améliorer à la fois le contrôle politique et la gestion locale. En effet les IA permettent de traiter une multitude de signaux issus de nombreux capteurs

implantés dans les villes et l'ensemble d'un pays afin de contrôler leurs citoyens et éviter de dépendre de sources humaines « peu fiables ».

Cette plus grande maîtrise de l'IA par les régimes autoritaires fait naître des craintes quant à la capacité des démocraties à pouvoir tirer parti de ces technologies pour protéger les libertés. Ben Buchanan, par exemple, insiste sur la nécessité pour les dirigeants des pays démocratiques de rester vigilants face à ces nouvelles menaces.

En matière de concurrence géopolitique, une possibilité inquiétante transparait dans cet ouvrage : l'IA favorisera plus les autocraties que les démocraties. Dans cette optique, les autocrates seront simplement plus capables ou plus disposés à utiliser l'IA pour contrôler les personnes, les informations et les armes. Libérés des lois sur la protection de la vie privée, ils auront accès à davantage de données. Ils utiliseront leurs systèmes de planification centralisée pour mener des recherches algorithmiques et construire des ordinateurs toujours plus rapides, transformant la plus grande faiblesse de l'autocratie – le

besoin de centralisation – en sa plus grande force. L'IA contribuera à ancrer le pouvoir au sein de l'État, en facilitant la surveillance et la répression et en obligeant les entreprises privées à apporter leur aide si nécessaire. Pour la sécurité nationale, l'IA améliorera également les outils de l'arsenal de l'autocrate, tels que les armes autonomes létales, les cyberopérations et la désinformation. Dans les autocraties, les préoccupations éthiques ne seront que des obstacles mineurs⁶².

Il existe une tension de plus en plus importante entre les objectifs commerciaux des plateformes dont les modèles économiques sont basés sur l'acquisition des nouveaux usagers (et donc de leurs données personnelles), et les préoccupations légitimes des pouvoirs publics dans les sociétés démocratiques pour limiter la diffusion des paroles de haine et des campagnes de désinformation. À ce titre la protection des démocraties devra aussi passer par une réflexion démocratique sur les modèles économiques de ces plateformes dont le rôle politique devrait s'accroître encore avec la montée en puissance des IA.

⁶² "When it comes to geopolitical competition, a worrying possibility runs throughout this book: that AI will do more for autocracy than it will for democracy. In this view, autocrats will simply be better able or more willing to use AI to control people, information, and weapons. Unencumbered by privacy laws, they will have access to more data. They will use their central planning systems to drive algorithmic research and build ever-faster computers, turning autocracy's biggest weakness—the need for centralization—into its greatest strength. AI will help entrench power within the state, aiding in surveillance and repression and compelling private companies to assist as needed. For national security, the argument goes, AI will also improve the tools in the autocrat's arsenal, such as lethal autonomous weapons, cyber operations, and disinformation. In autocracies, ethical concerns will be nothing but minor speed bumps."

Ben Buchanan, Andrew Imbrie. *The New Fire: War, Peace, and Democracy in the Age of AI* (p. 8) (MIT Press 2022)

D. Les vulnérabilités des démocraties face à l'IA

L'ouverture des sociétés démocratiques aux influences étrangères a souvent été décrite par des régimes hostiles comme une vulnérabilité politique et idéologique. Désormais, les usages des technologies créent de nouvelles surfaces d'attaque sur nos infrastructures et sur le fonctionnement social et politique de nos sociétés. Ainsi, avec la montée en puissance des technologies d'IA, la dynamique des conflits pourrait être profondément modifiée au profit des cyber-attaquants. Ce que les auteurs de *L'Âge de l'IA* résument ainsi :

À mesure que les économies avancées intègrent des systèmes de commandement et de contrôle numériques dans les centrales électriques et les réseaux électriques, transfèrent leurs programmes gouvernementaux sur de grands serveurs et des systèmes cloud et transfèrent les données dans des registres électroniques, leur vulnérabilité aux cyberattaques se multiplie ; elles représentent alors un ensemble de cibles plus riche, une attaque réussie pourrait alors être considérablement dévastatrice. À l'inverse, en cas de perturbation numérique, les États à faible niveau technologique, les organisations terroristes et même les attaquants individuels pourraient estimer qu'ils ont beaucoup moins à perdre⁶³.

Les limites du rôle « émancipateur » des technologies numériques ont été étudiées depuis le début de l'Internet grand public. Dès 2001, les chercheurs du *Carnegie Endowment for International Peace* évoquaient leurs doutes quant à la capacité des démocraties à mieux maîtriser ces technologies que les régimes autoritaires :

⁶³ " As advanced economies integrate digital command-and-control systems into power plants and electricity grids, shift their governmental programs onto large servers and cloud systems, and transfer data into electronic ledgers, their vulnerability to cyberattack multiplies; they present a richer set of targets so that a successful attack could be substantially devastating. Conversely, in the event of a digital disruption, the low-tech state, the terrorist organization, and even individual attackers may assess that they have relatively much less to lose. "

Henry A. Kissinger, Eric Schmidt and Daniel Huttenlocher. *The Age of AI: And Our Human Future*. (John Murray Press, 2021)

Les régimes autoritaires devront adapter en permanence leurs dispositifs de contrôle s'ils veulent relever efficacement les défis posés par les futures évolutions des technologies de l'information et de la communication. Il est vraiment possible que cette tâche s'avère trop difficile et que l'utilisation des technologies de l'information finisse par jouer un rôle dans la révolution démocratique si souvent annoncée. Au fil du temps, cependant, les régimes autoritaires ont surmonté d'innombrables défis posés par l'évolution des technologies et ils pourraient bien se montrer à la hauteur du défi actuel⁶⁴.

Le réseau X (ex-Twitter) est souvent décrit comme le réseau social de la colère et de l'instantanéité, propice à la polarisation des opinions au sein de bulles informationnelles. Cette tendance s'est encore renforcée depuis qu'Elon Musk a entrepris d'y déployer sa vision d'une liberté d'expression sans aucune limite. Cela a rendu la plateforme X encore plus accessible aux opérations de déstabilisation et d'ingérence menées par des États hostiles ainsi que par des groupes politiques ou religieux radicaux. Avec le développement des IA génératives, nous pourrions connaître une étape encore plus favorable à la manipulation des opinions. En effet, les IA génératives permettent de modéliser les opinions individuelles afin qu'elles répondent aux sollicitations de la bulle informationnelle ultime : celle qui serait construite entre chaque usager et des IA, sans plus aucun intermédiaire humain.

⁶⁴ Authoritarian regimes will have to continually adapt their measures of control if they want to counter effectively the challenges of future variations in information and communication technologies. It is quite possible that this task will prove too difficult and that use of ICTs will eventually play a role in the democratic revolution that has been so widely predicted. Over time, however, authoritarian regimes have weathered innumerable challenges posed by changing technologies, and they may prove up to the current challenge as well. The Internet and State Control in Authoritarian Regimes: China, Cuba, and the Counterrevolution Rapport du Carnegie Endowment for International Peace par Shanthi Kalathil et Taylor C. Boas (CEIP 2001) <https://carnegie-production-assets.s3.amazonaws.com/static/files/21KalathilBoas.pdf>

E. Chine : l'IA au service du contrôle politique des populations

Depuis lors, de nombreux régimes autoritaires ont démontré qu'ils pouvaient limiter les risques de déstabilisation interne et dans le même temps bénéficier de ces technologies à des fins de contrôle idéologique des populations. La Chine a ainsi dans un premier temps élaboré son grand firewall (*Great Firewall of China*) afin de contrôler et censurer l'ensemble des flux d'informations entrant et sortant

“ *En Chine, le Crédit social permet d'accroître les connaissances sur les individus et les entreprises et d'améliorer les algorithmes d'IA qui permettront à leur tour d'analyser plus finement les données permettant de mieux contrôler les populations...*

de son territoire afin d'éviter la contagion idéologique issue des pays démocratiques. Par la suite, les autorités chinoises ont développé un dispositif de surveillance de masse sans équivalent dans le monde : le *Crédit social* (ou *Social Credit System*). Ce dispositif orwellien élaboré par certaines des plus puissantes sociétés chinoises comme *Alibaba*, permet d'attribuer une note à l'ensemble des citoyens chinois. Le *Crédit social* repose à la fois sur des technologies de reconnaissance et des algorithmes d'IA conçus pour analyser le comportement des individus. La note affectée à chaque citoyen chinois a pour

objectif d'évaluer son « comportement » social, financier et idéologique⁶⁵. Une note de *Crédit social* trop basse leur interdit de se déplacer en train ou en avion, d'accéder à certains services publics ou encore d'obtenir un crédit. Pour les concepteurs du *Crédit social*, il s'agit d'éliminer du corps social les éléments potentiellement perturbateurs. Selon Hou Yunchun, l'ancien directeur adjoint du centre de recherche sur le développement du *Conseil d'État*, le but du *Crédit Social*

⁶⁵ China's Social Credit System (Infographie de la Bertelsmann Foundation 2019)

www.bertelsmann-stiftung.de/fileadmin/files/aam/Asia-Book_A_03_China_Social_Credit_System.pdf

est simplement de « *mettre les personnes discréditées en faillite...*⁶⁶ ». Le régime chinois a ainsi démontré avec le *Crédit social* qu'il pouvait créer un système de surveillance et d'éradication de la dissidence à l'échelle de sa population tout entière. De même qu'il existait une compétition entre les démocraties sur l'usage des technologies à des fins économiques, une autre forme d'émulation s'est créée entre régimes autoritaires pour améliorer les technologies de surveillance de leurs citoyens *via* les IA : depuis le suivi de la géolocalisation en temps réel, la reconnaissance faciale, jusqu'à l'analyse des comportements et des activités des citoyens. Dans le même temps ces technologies ont permis à des puissances étrangères d'influencer les opinions des démocraties. Les autocraties ont ainsi pris appui sur les expériences de psychologie sociale qui avaient été menées par les réseaux sociaux pour manipuler les émotions de leurs usagers⁶⁷. Sur le plan politique, il devenait possible de connaître l'impact des campagnes politiques dans les populations et aussi développer de nouveaux savoir-faire dans le domaine de la désinformation. C'est d'ailleurs en s'appuyant sur les connaissances des experts russes spécialisés dans la désinformation que les équipes de la société *Cambridge Analytica* ont, en 2016, mis en place leur campagne visant à favoriser l'élection de Donald Trump⁶⁸.

Dans son ouvrage *Algorithmes : la bombe à retardement*, la mathématicienne Cathy O'Neil avait décrit, deux ans avant les révélations de l'affaire *Cambridge Analytica*, les mécanismes de « microciblage » qui peuvent être utilisés pour manipuler les électeurs indécis : « *Le marketing moderne offre ainsi aux politiciens de nouveaux moyens d'atteindre des électeurs spécifiques afin de leur dire ce qu'ils veulent entendre. Une fois qu'ils l'auront fait, ces électeurs seront susceptibles d'accepter*

⁶⁶ China blacklists millions of people from booking flights as 'social credit' system introduced (The Independent, 22 Nov 2018) www.independent.co.uk/news/world/asia/china-social-credit-system-flight-booking-blacklisted-beijing-points-a8646316.html

⁶⁷ Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock : Experimental evidence of massive-scale emotional contagion through social networks (Proceedings of the National Academy of Sciences PNAS 2 Jun 2014) www.pnas.org/doi/10.1073/pnas.1320040111

⁶⁸ Cambridge Analytica: links to Moscow oil firm and St Petersburg university (The Guardian, 17 Mar 2018) www.theguardian.com/news/2018/mar/17/cambridge-academic-trawling-facebook-had-links-to-russian-university

l'information pour argent comptant, car elle confirme leurs idées reçues ou leurs préjugés, un phénomène que les psychologues appellent le biais de confirmation⁶⁹. »

Ainsi, comme le précise la *MIT Tech Review*, les exemples successifs de négligence de la part des principales plateformes de l'Internet ont contraint les nouveaux acteurs de l'IA à mieux prendre en compte la dimension politique de leurs activités, en particulier lorsqu'il est question de désinformation et d'ingérence.

Les chercheurs s'attendaient depuis longtemps à ce que les technologies d'IA génératives, et en particulier les grands modèles de langage, soient adoptées par des acteurs hostiles, pour augmenter à moindre coût l'ampleur et l'efficacité de leurs opérations. La divulgation par la société OpenAI de manière transparente du fait que cela s'est produit – et qu'elle a donné la priorité à la détection de ce phénomène et à la fermeture de comptes pour atténuer son impact – montre qu'au moins une grande entreprise d'IA a tiré des enseignements des difficultés des médias sociaux dans les années qui ont suivi l'ingérence de la Russie dans l'élection américaine de 2016...⁷⁰

La Chine a longtemps utilisé sa puissance technologique à des fins de promotion de son régime et des bienfaits d'un « capitalisme contrôlé » par le *Parti Communiste Chinois*, puis a dans un second temps utilisé ses instruments technologiques pour influencer les débats politiques en Europe et aux États-Unis. Dans ce domaine, la Chine a longuement analysé les procédés mis en place par la Russie dans ses

⁶⁹ Modern consumer marketing, however, provides politicians with new pathways to specific voters so that they can tell them what they know they want to hear. Once they do, those voters are likely to accept the information at face value because it confirms their previous beliefs, a phenomenon psychologists call confirmation bias. Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* by Cathy O'Neil (p. 187). (Crown 2016)

⁷⁰ Researchers have long expected adversarial actors to adopt generative AI technology, particularly large language models, to cheaply increase the scale and caliber of their efforts. The transparent disclosure that this has begun to happen—and that OpenAI has prioritized detecting it and shutting down accounts to mitigate its impact—shows that at least one large AI company has learned something from the struggles of social media platforms in the years following Russia's interference in the 2016 US election.

Propagandists are using AI too—and companies need to be open about it (MIT Technology Review 8 Jun 2024) www.technologyreview.com/2024/06/08/1093356/propagandists-are-using-ai-too-and-companies-need-to-be-open-about-it/

campagnes de désinformation. Comme le montre le rapport de Paul Charon & Jean-Baptiste Jeangène Vilmer : *Les opérations d'influence chinoises : un moment machiavélien* publié par l'Institut de Recherche Stratégique de l'École Militaire (IRSEM).

Chine : pour l'Armée Populaire de Libération (APL) la Russie est un modèle à imiter⁷¹

En matière de manipulation des réseaux sociaux, la Russie constitue pour la Chine « un modèle à imiter », depuis l'annexion de la Crimée en 2014, qui a été suivie de près en Chine, particulièrement au sein de l'APL. En 2014, un membre du département général de l'état-major de l'APL a ainsi écrit un article tirant trois leçons de la « guerre de l'opinion publique » russe en Ukraine : « prenez l'offensive en poussant votre récit en premier, présentez vos arguments juridiques et soutenez le tout avec du hard power ». Plusieurs autres articles similaires ont paru les années suivantes, témoignant d'une claire volonté des militaires chinois d'apprendre de l'exemple russe. En 2018, un article étudiant la manière dont RT (anciennement *Russia Today*) a couvert les frappes américaines en Syrie recommandait d'étudier la méthode de communication de RT sans perdre « l'objectivité », « nous pouvons influencer silencieusement les émotions et les inclinations du public et le rendre dépendant des informations de nos médias ». RT est régulièrement citée comme un modèle à suivre dans la littérature militaire chinoise, en particulier pour son activité sur les réseaux sociaux : des analystes de l'Université de défense nationale comparent la chaîne russe à « un porte-avions de propagande », soulignant notamment sa performance sur *YouTube*.

⁷¹ www.irsem.fr/rapport.html

Désormais, avec *TikTok*, sa plateforme de partage de vidéos qui compte plus de 1,5 milliard d'utilisateurs, la Chine dispose de l'un des plus importants canaux d'influence jamais créé en direction des populations occidentales. Comme les démocraties ont pu le constater, le régime chinois utilise ses outils pour recueillir des informations sensibles, qu'elles soient idéologiques, religieuses ou médicales sur les utilisateurs de sa plateforme pour être en mesure de les influencer. Ainsi dans leur rapport au Sénat : « *La tactique TikTok : opacité, addiction et ombres chinoises* », MM. Mickaël Vallet et Claude Malhuret notaient à propos des risques que représente cette plateforme :

« Force de frappe internationale au succès incontestable, TikTok pourrait être utilement mis au service des intérêts et de la pérennité du régime chinois, au détriment de la sécurité des données des utilisateurs mondiaux de TikTok, du pluralisme des expressions et de la qualité de l'information diffusée sur sa plateforme. Comme le soulignait Bernard Benhamou, secrétaire général de l'Institut de la souveraineté numérique : « Par définition, ne pas utiliser ces données serait une forme de faute professionnelle de la part des services chinois⁷² ».

Désormais, au-delà de la seule propagande visant à promouvoir le régime chinois, il est question d'ingérence dans les principaux processus électoraux des démocraties. Récemment, les services chinois ont emprunté aux services russes leurs méthodologies en matière de désinformation pour développer leurs propres actions d'ingérence dans le processus électoral américain⁷³. Il leur devient possible d'amplifier artificiellement des messages auprès de milliers voire de millions de profils en ciblant spécifiquement des personnes ayant montré une sensibilité sur certains sujets. Cette amplification algorithmique peut se dérouler sans même que l'émetteur du message originel en ait connaissance.

⁷² Extrait du rapport du Sénat "La tactique TikTok : opacité, addiction et ombres chinoises" - n° 831 (2022-2023) déposé le 4 juillet 2023 www.senat.fr/rap/r22-831-1/r22-831-11.pdf

⁷³ China's Advancing Efforts to Influence the U.S. Election Raise Alarms (New York Times, 2 Apr 2024) www.nytimes.com/2024/04/01/business/media/china-online-disinformation-us-election.html

Pour des pays comme la Russie ou la Chine, les actions de la guerre hybride ne constituent pas seulement un moyen de limiter les risques de contagion démocratique mais aussi de limiter les capacités d'innovations des pays adverses. Ces capacités menacent en effet leurs régimes tant sur le plan économique et militaire que sur le plan politique par l'attractivité qu'elles peuvent représenter pour leurs élites scientifiques.

VI. TRANSHUMANISME, LONG-TERMISME, EUGÉNISME : DES IDÉOLOGIES AU SERVICE D'UNE VISION RADICALE DE L'HUMAIN

Les technologies peuvent être inspirées par les conceptions politiques philosophiques ou religieuses de leurs concepteurs et peuvent à leur tour transmettre ces conceptions à leurs utilisateurs. Ainsi, en 1994, le philosophe Umberto Eco décrivait avec humour ce qu'il considérait être les racines religieuses de l'affrontement entre *Microsoft* et *Apple*, en étudiant les différences entre les premiers systèmes d'exploitation *MS-DOS* et *Macintosh* :

« Le fait est que le monde est divisé entre les utilisateurs d'ordinateurs Macintosh et les utilisateurs d'ordinateurs compatibles MS-DOS. Je suis fermement convaincu que le Macintosh est catholique et que le DOS est protestant. En effet, le Macintosh est contre-réformiste et a été influencé par le ratio studiorum des Jésuites⁷⁴. C'est gai, amical, conciliant ; il indique aux fidèles comment ils doivent procéder étape par étape pour atteindre – sinon le royaume des Cieux – le moment où leur document sera imprimé. C'est un catéchisme : l'essence de la révélation est traitée à travers des formules simples et des icônes somptueuses. Tout le monde a droit au salut. MS-DOS est protestant, voire calviniste. Il permet une interprétation libre des Écritures, exige des décisions personnelles difficiles, impose une herméneutique subtile à l'utilisateur et tient pour acquis l'idée que tous ne peuvent pas parvenir au salut. Pour que le système fonctionne, il faut interpréter soi-même le programme : loin de la communauté baroque des fêtards, l'utilisateur est enfermé dans la solitude de son propre tourment intérieur⁷⁵. »

⁷⁴ La raison des études. Sens et histoire du Ratio Studiorum (Georges Leroux 1995- Revue Erudit 2024) www.erudit.org/fr/revues/etudfr/1995-v31-n2-etudfr1079/035976ar.pdf

⁷⁵ The fact is that the world is divided between users of the Macintosh computer and users of MS-DOS compatible computers. I am firmly of the opinion that the Macintosh is Catholic and that DOS is Protestant. Indeed, the Macintosh is counter-reformist and has been influenced by the ratio studiorum of the Jesuits. It is cheerful, friendly, conciliatory; it tells the faithful how they must proceed step by step to reach -- if not the kingdom of Heaven -- the moment in which their document is printed. It is catechistic: The essence of revelation is dealt with via simple formulae and sumptuous icons. Everyone has a right to salvation.

À l'inverse aujourd'hui ce sont les technologies qui sont à l'origine de nouvelles idéologies qui visent à prendre en compte leurs effets sur les sociétés voire sur l'espèce humaine tout entière. À l'extrême, les IA pourraient à l'avenir être à l'origine de nouveaux cultes qui s'appuieront sur les avis émis par l'IA. Ces avis pourraient être perçus comme relevant d'une transcendance auxquelles seules des IA auraient accès. Pour certains philosophes, les êtres humains tentés par l'abandon à l'Intelligence artificielle, expriment une forme moderne de « fatigue de soi ». C'est le cas des transhumanistes qui développent une croyance en une version idéalisée de l'humanité qui emprunterait la voie de la machine pour s'extraire de sa condition (faillibilité, mortalité, vieillissement...). Ainsi, pour Susan Schneider : « *Certains imaginent en effet que l'intelligence synthétique sera la prochaine étape de l'évolution de l'intelligence sur Terre. Vous et moi, la façon dont nous vivons et expérimentons le monde aujourd'hui, ne serions qu'une étape vers l'IA, un barreau sur l'échelle de l'évolution* ⁷⁶ ».

A. Transhumanisme : vers des humains inhumains ?

Pour les industriels des technologies, le credo transhumaniste s'affirme commodément comme un moyen de repousser toutes les régulations éthiques des technologies au nom d'un lointain intérêt supérieur de l'espèce humaine. Cela s'effectue aux dépens de la réflexion sur les risques immédiats que ces innovations pourraient représenter pour les sociétés. À titre d'exemple, pour les tenants de la philosophie « long-termiste », la crise climatique ne serait qu'un épiphénomène

MS-DOS is Protestant, or even Calvinistic. It allows free interpretation of scripture, demands difficult personal decisions, imposes a subtle hermeneutics upon the user, and takes for granted the idea that not all can achieve salvation. To make the system work you need to interpret the program yourself: Far away from the baroque community of revelers, the user is closed within the loneliness of his own inner torment.

The Holy War: Mac vs. DOS English translation of Umberto Eco's back-page column, "La bustina di Minerva," in the Italian newsweekly Espresso (September 30, 1994).

www.agonia.net/index.php/prose/124266/The_Holy_War:_Mac_vs._DOS

⁷⁶ « Indeed, some suspect that synthetic intelligence will be the next phase in the evolution of intelligence on Earth. You and I, how we live and experience the world right now, are just an intermediate step to AI, a rung on the evolutionary ladder. »

Susan Schneider: Artificial You: AI and the Future of Your Mind (Princeton University Press 2019)

puisqu'elle ne pourrait pas à elle seule remettre en cause l'existence même de l'espèce humaine. Seule la destinée « multiplanétaire » de l'humanité mériterait d'être préservée, y compris au prix de la liberté des citoyens de nos sociétés actuelles. Pour le philosophe transhumaniste Nick Bostrom, il pourrait être nécessaire d'établir une surveillance mondiale au nom de l'intérêt supérieur de l'espèce humaine. Cette surveillance aurait pour but d'éviter que les actions des humains ne puissent les empêcher de réaliser leur destinée qui, pour lui, consisterait à permettre à l'espèce humaine de devenir multiplanétaire.

« Le lien avec le long-termisme est que, selon Bostrom et Ord, ne pas devenir posthumain nous empêcherait en apparence de réaliser notre vaste et glorieux potentiel, ce qui serait une catastrophe existentielle. Comme l'a dit Bostrom en 2012, « l'exclusion permanente de toute possibilité de ce type de changement transformateur de la nature biologique humaine pourrait en soi constituer une catastrophe existentielle ». De même, Ord affirme que « préserver à jamais l'humanité telle qu'elle est aujourd'hui pourrait également gaspiller notre héritage, en renonçant à la plus grande partie de notre potentiel ». Bostrom a lui-même soutenu que nous devrions sérieusement envisager d'établir un système de surveillance mondial et intrusif qui surveillerait en temps réel chaque personne sur la planète, afin d'amplifier les « capacités de police préventive »⁷⁷ (par exemple, pour prévenir les attaques terroristes « omnicidaires » qui pourraient dévaster la civilisation)⁷⁸. »

Cette vision apocalyptique du futur imaginé par Bostrom apparaîtrait anecdotique si elle n'avait été accueillie favorablement par les plus puissants patrons libertariens

⁷⁷ Bostrom, Nick (2013). Existential Risk Prevention as Global Priority. Global Policy, 4 (1), 15–31.
<https://existential-risk.com/concept.pdf>

⁷⁸ The connection with longtermism is that, according to Bostrom and Ord, failing to become posthuman would seemingly prevent us from realising our vast and glorious potential, which would be existentially catastrophic. As Bostrom put it in 2012, 'the permanent foreclosure of any possibility of this kind of transformative change of human biological nature may itself constitute an existential catastrophe.' Similarly, Ord asserts that 'forever preserving humanity as it is now may also squander our legacy, relinquishing the greater part of our potential.' Bostrom himself argued that we should seriously consider establishing a global, invasive surveillance system that monitors every person on the planet in realtime, to amplify the 'capacities for preventive policing' (eg, to prevent omnicidal terrorist attacks that could devastate civilisation).

Why longtermism is the world's most dangerous secular credo (Émile P Torres in Aeon Essays 19 Oct 2021)
<https://aeon.co/essays/why-longtermism-is-the-worlds-most-dangerous-secular-credo>

de la Silicon Valley ; depuis Elon Musk ou Peter Thiel jusqu'à Jeff Bezos ou encore Samuel Bankman-Fried, le magnat déchu des cryptomonnaies.

Ces opinions peuvent aussi se refléter dans la conception des technologies et en particulier des modèles d'IA. Ainsi, Elon Musk a tenté, jusqu'ici avec un succès relatif⁷⁹, de créer une AI générative qui soutient sa vision libertarienne et qui constitue le pendant conservateur à des IA comme *ChatGPT* ou *Gemini* développées par *OpenAI* et *Google*.

Extrait de la Déclaration Transhumaniste⁸⁰

- L'avenir de l'humanité va être radicalement transformé par la technologie. Nous envisageons la possibilité que l'être humain puisse subir des modifications, telle que son rajeunissement, l'accroissement de son intelligence par des moyens biologiques ou artificiels, la capacité de moduler son propre état psychologique, l'abolition de la souffrance et l'exploration de l'univers.
- On devrait mener des recherches méthodiques pour comprendre ces futurs changements ainsi que leurs conséquences à long terme.
- Les transhumanistes croient que, en étant généralement ouverts à l'égard des nouvelles technologies, et en les adoptant au lieu d'essayer de les interdire nous favoriserions leur utilisation à bon escient.

⁷⁹ I tried X's 'anti-woke' Grok AI chatbot. The results were the opposite of what I expected (ZDNET Dec. 21,2023) www.zdnet.com/article/i-tried-xs-anti-woke-grok-ai-chatbot-the-results-were-the-opposite-of-what-i-expected/

⁸⁰ The Transhumanist Declaration;

1 Humanity will be radically changed by technology in the future. We foresee the feasibility of redesigning the human condition, including such parameters as the inevitability of aging, limitations on human and artificial intellects, unchosen psychology, suffering, and our confinement to the planet earth.

2 Systematic research should be put into understanding these coming developments and their long-term consequences.

3 Transhumanists think that by being generally open and embracing of new technology we have a better chance of turning it to our advantage than if we try to ban or prohibit it. »

Susan Schneider Artificial You: AI and the Future of Your Mind (Princeton University Press 2019)

En dehors des sources utilisées pour entraîner ces IA génératives, qui peuvent avoir un impact crucial sur les réponses fournies par ces IA, leurs ingénieurs peuvent aussi y introduire, avec des risques évidents de biais ou de discrimination, leurs propres priorités morales. Ainsi, Brian Christian rappelait que pour guider leurs travaux, certains concepteurs de modèles d'IA pouvaient même remettre en cause le bien-fondé des valeurs morales de nos sociétés.

« Certains, par exemple, s'inquiètent même du fait que les humains ne constituent pas une référence en matière d'autorité morale : « Nous avons beaucoup parlé du problème de l'introduction de valeurs humaines dans les machines », déclare Blaise Agüera y Arcas, de Google. « En fait, je ne pense pas que ce soit le principal problème. Je pense que le problème est que, telles qu'elles existent, les valeurs humaines ne suffisent pas. Elles ne sont pas assez bonnes. ⁸¹ ».

À l'inverse, Jaron Lanier, l'un des pionniers de la réalité virtuelle, insiste sur les dangers politiques et sociaux que représentent les courants de pensée comme le transhumanisme et le « long-termisme ». Selon lui en effet, ces idées correspondent à une nouvelle forme de déterminisme technologique qui comporte des risques sociaux et culturels majeurs : *« Je ne pense pas que la technologie se crée d'elle-même. Ce n'est pas un processus autonome. La raison pour laquelle nous croyons en l'action humaine plutôt qu'au déterminisme technologique est que nous pouvons alors avoir une économie dans laquelle les gens gagnent leur vie et inventent leur propre vie. Si vous structurez une société sans mettre l'accent sur l'action humaine individuelle, cela revient pratiquement à priver les individus de leur influence, de leur dignité et de leur autodétermination ⁸².*

⁸¹ Some, for instance, worry that humans aren't a particularly good source of moral authority. "We've talked a lot about the problem of infusing human values into machines," says Google's Blaise Agüera y Arcas. "I actually don't think that that's the main problem. I think that the problem is that human values as they stand don't cut it. They're not good enough."

Brian Christian. The Alignment Problem, Machine Learning and Human Values (p. 247)(W. W. Norton & Company 2020)

⁸² "I do not think the technology is creating itself. It's not an autonomous process. The reason to believe in human agency over technological determinism is that you can then have an economy where people earn their

B. IA et transformations culturelles des sociétés

Les IA génératives commencent déjà à transformer certaines professions créatives et pourraient en menacer d'autres, comme en témoignent, aux États-Unis, les mouvements sociaux récents des scénaristes, des doubleurs ou des acteurs inquiets d'être remplacés par des IA génératives dans les temps à venir. Pour Yuval Harari, l'étape suivante pourrait être des créations culturelles entièrement « autonomes » conçues par des IA qui viendront coloniser les esprits et modifier les pensées et les rêves des humains. Là où les créations artistiques témoignaient des expériences, de la culture et des valeurs de leurs auteurs, ces nouvelles créations pourraient même être à l'origine de nouveaux dogmes, voire de nouveaux cultes, qui pourraient transformer de manière imprévisible l'imaginaire des humains et le comportement des sociétés.

Qu'advient-il du cours de l'histoire lorsque l'IA s'emparera de la culture et commencera à produire des histoires, des mélodies, des lois et des religions ? Des outils antérieurs tels que l'imprimerie et la radio ont aidé à diffuser les idées culturelles des humains, mais n'ont jamais créé leurs propres idées culturelles. L'IA est fondamentalement différente, elle peut créer des idées complètement nouvelles, une toute nouvelle culture. Au début, l'intelligence artificielle va probablement imiter les prototypes humains avec lesquels elle s'est formée à ses débuts. Mais au fil des ans, la culture de l'intelligence artificielle osera arriver là où aucun être humain n'est arrivé auparavant. Pendant des millénaires, les êtres humains ont vécu dans les rêves d'autres êtres humains. Au cours des prochaines décennies, nous pourrions nous retrouver à vivre dans les rêves d'une intelligence extraterrestre.⁸³

own way and invent their own lives. If you structure a society on not emphasizing individual human agency, it's the same thing operationally as denying people clout, dignity, and self-determination..."

Lanier, Jaron : "Who Owns the Future?" (Simon & Schuster 2013)

⁸³ What will happen to the course of history when ai takes over culture, and begins producing stories, melodies, laws and religions? Previous tools like the printing press and radio helped spread the cultural ideas of humans, but they never created new cultural ideas of their own. ai is fundamentally different. ai can create completely new ideas, completely new culture. At first, ai will probably imitate the human prototypes that it was trained on in its infancy. But with each passing year, ai culture will boldly go where no human has gone before. For

De nombreuses normes et habitudes sociales ont été modifiées du fait de l'essor des technologies numériques. Les réseaux sociaux ont ainsi transformé notre rapport à l'exposition des données personnelles. Des dispositifs comme les lunettes de réalité augmentée pourraient constituer d'autres ruptures dans les normes interpersonnelles. La possibilité de captation vidéo et audio permanente pourrait devenir la nouvelle norme lors des échanges et des conversations. Pour les acteurs des technologies dont l'activité repose sur l'utilisation massive des données personnelles à des fins publicitaires, toute entrave à la collection de données devient non seulement un obstacle économique, mais aussi une nuisance politique. Ce qui faisait dire à Eric Schmidt, alors patron de *Google*, qu'il était nécessaire de faire évoluer notre conception de la vie privée plutôt que de tenter de la préserver⁸⁴. En effet, certains acteurs industriels considèrent qu'il est préférable de modifier les normes sociales ou culturelles de nos sociétés, afin que puissent se développer leurs technologies. À l'extrême, il conviendrait non seulement de modifier ces normes sociales ou culturelles mais bien le patrimoine génétique qui constitue les humains eux-mêmes, pour les rendre « compatibles » avec certaines évolutions technologiques.

millennia human beings have lived inside the dreams of other humans. In the coming decades we might find ourselves living inside the dreams of an alien intelligence.

Yuval Noah Harari argues that AI has hacked the operating system of human civilisation (The Economist 28 Apr 2023) www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation

⁸⁴ Google CEO Eric Schmidt Dismisses the Importance of Privacy (Electronic Frontier Foundation 10 Dec 2009) www.eff.org/fr/deeplinks/2009/12/google-ceo-eric-schmidt-dismisses-privacy

C. Du transhumanisme à l'eugénisme

« Vos capacités m'intriguent Capitaine, mais honnêtement vous restez mentalement et physiquement inférieurs. En fait, je suis surpris du peu d'améliorations apportées par l'évolution humaine [...] Rien ne change jamais, sauf l'homme. Vos réalisations techniques ? Améliorez un appareil mécanique et vous pouvez doubler la productivité. Mais améliorez l'homme et vous y gagnez par un facteur 1000. Je suis un tel homme. Rejoignez-moi, je vous traiterai bien... »

Khan Noonien Singh, tyran génétiquement « augmenté » s'adressant au capitaine Kirk dans l'épisode *Space Seed* de *Star Trek* (1967)⁸⁵

De nouveaux services pourraient être rendus accessibles par des modifications qui ne seront pas liées aux évolutions des codes sociaux ou culturels mais bien à des modifications du code biologique lui-même. En dehors de leurs applications médicales, les avancées liées à la génomique pourraient ainsi être mises à contribution pour créer de nouvelles générations de services qui « augmenteront » les capacités des êtres humains. De nouvelles générations de services pourraient être créées qui prendraient racine non plus dans des interfaces homme-machine traditionnelles ou en réalité virtuelle mais directement dans des connexions avec le cerveau des utilisateurs. C'est l'objectif de sociétés spécialisées dans les neurotechnologies, comme *Neuralink* créée par Elon Musk en 2016 : développer des implants cérébraux d'interfaces directes dites neuronales. Plus récemment, la start-up *Merge Labs* financée par le patron d'*OpenAI* Sam Altman⁸⁶ envisage d'utiliser des systèmes moléculaires pour interagir avec les neurones humains. Dans un premier temps, ces technologies devraient permettre de rétablir de

⁸⁵ Captain, although your abilities intrigue me, you are quite honestly inferior. Mentally, physically. In fact, I am surprised how little improvement there has been in human evolution [...] Nothing ever changes, except man. Your technical accomplishments? Improve a mechanical device and you may double productivity. But improve man and you gain a thousandfold. I am such a man. Join me. I'll treat you well...

Khan Noonien Singh, genetically augmented tyrant addressing Captain Kirk in Star Trek episode Space Seed (1967) <http://www.chakoteya.net/StarTrek/24.htm>

⁸⁶ OpenAI Invests in Sam Altman's New Brain-Tech Startup Merge Labs (Wired, 15 Jan 2026) www.wired.com/story/openai-invests-in-sam-altmans-new-brain-tech-startup-merge-labs/

nouvelles formes de mobilité et ou d'interaction pour des personnes handicapées. Par la suite, des services liés au bien-être, voire des « extensions » mémorielles ou des modifications de la conscience pourraient être créés pour des humains dits « augmentés ». Ainsi, le neuroscientifique chinois Mu-ming Poo, de *l'Institut de neuroscience de Shanghai*, met en garde sur les risques de dérives des neurotechnologies et sur les « *questions éthiques plus difficiles [liées à] leurs utilisations pour augmenter les fonctions cérébrales normales plutôt qu'à des fins*

“ Il ne s'agit pas de savoir quand l'IA deviendra égale ou supérieure à l'homme mais bien quand l'homme « augmenté » cessera d'être humain... À partir de quand des modifications génétiques [et/ou ajouts de prothèses cérébrales] transformeront les êtres qui s'y livreront en des êtres « non-humains »...

thérapeutiques, [pour] des applications telles que la neuromodulation non invasive du sommeil et de l'éveil, de l'attention, de l'émotion, ainsi que des fonctions d'apprentissage et de la mémoire⁸⁷ ».

Pour arriver à cette fusion entre l'homme et la machine, certains imaginent déjà des modifications génétiques qui rendraient les humains plus aptes à cette connexion avec des systèmes technologiques⁸⁸. Il existerait alors une continuité entre les concepts transhumanistes et l'eugénisme qui permettrait, selon leurs promoteurs, d'augmenter génétiquement les êtres humains pour qu'ils soient à la fois plus

« performants » et mieux à même d'interagir avec l'univers technologique. La question qui se poserait alors serait l'exact inverse de celle qui occupe les spécialistes de l'IA générale : il ne s'agit pas de savoir quand l'IA deviendra égale ou supérieure à l'homme mais bien quand l'homme cessera d'être humain une fois « augmenté ». Et à partir de quel moment ces modifications génétiques et/ou

⁸⁷ Comment lire et stimuler le cerveau devient accessible à tous... pour le meilleur et pour le pire (Le Monde, 28 oct 2024) www.lemonde.fr/sciences/article/2024/10/28/du-soin-a-la-surveillance-du-cerveau-sept-defis-pour-les-neurotechnologies_6363498_1650684.html

⁸⁸ Les neurotechnologies : défis scientifiques et éthiques (Office parlementaire d'évaluation des choix scientifiques et technologiques - janvier 2022) www.senat.fr/fileadmin/Office_et_delegations/OPECST/Notes_scientifiques/OPECST_note32.pdf

ajouts de prothèses cérébrales transformeront les êtres qui s'y livreront en des êtres « non-humains ». Quelles seraient les capacités distinctives des humains qu'ils perdraient en ayant un cerveau connecté ? Quelles caractéristiques cérébrales seraient altérées par des manipulations génétiques ? Quelles fonctions actuelles de nos cerveaux seraient jugées inutiles par les auteurs de ces modifications génétiques ou technologiques ?

Paradoxalement, c'est l'essor conjoint de l'intelligence artificielle et des biotechnologies qui pourrait contraindre nos sociétés à redéfinir les limites entre ce que nous considérons comme socialement et biologiquement acceptable et ce qui ne pourrait pas l'être pour le devenir de l'espèce humaine. Des débats démocratiques de même nature que ceux qui ont eu lieu sur la bioéthique devront aussi exister pour déterminer quelles pourront être les modifications permises et celle qui, comme pour les IA, comporteront des risques inacceptables pour chaque individu et pour nos sociétés tout entières.

VII. PERSPECTIVES TECHNOLOGIQUES DES IA ET ÉVOLUTIONS POLITIQUES

A. Des IA encore éloignées de « *l'intelligence artificielle générale* »

Lorsque les IA génératives présentent des affirmations qui semblent certaines mais qui sont en réalité erronées, leurs concepteurs évoquent souvent le terme « d'hallucination ». Cela s'explique par le fait que ces systèmes, manquent de sens commun et de compréhension du contexte dans lequel des événements se produisent. Ces IA ne peuvent pas différencier les informations factuelles des affirmations incorrectes. Malgré les efforts pour perfectionner leurs modèles et affiner les données utilisées pour leur entraînement, ces IA peuvent produire des réponses qui contredisent des faits évidents. En un sens, toutes les réponses produites par des IA ne sont que des hallucinations, qui sont jugées plus ou moins utiles ou acceptables par ceux qui reçoivent ces réponses.

Ainsi, les chercheurs en intelligence artificielle continuent de débattre de la possibilité que des « *IA générales* » (ou *AGI*, pour *Artificial General Intelligence*) ou même des *ASI* (*Artificial Super Intelligence*) puissent voir le jour. Pour certains d'entre eux, des limites fondamentales des technologies d'IA pourraient rendre cet objectif inaccessible. Des débats de même nature existent sur la possibilité de concevoir de véritables ordinateurs quantiques qui pourraient transformer à la fois le paysage informatique mondial et celui de l'IA. Ainsi Alain Aspect, prix Nobel de physique 2022 et spécialiste mondial de l'intrication quantique, n'excluait pas que puissent exister des obstacles fondamentaux qui interdiraient la mise en œuvre de ces ordinateurs quantiques :

L'ordinateur quantique « idéal », qui est celui dont on entend parler, n'existe pas encore. Personnellement, je pense que c'est tellement difficile à réaliser que je doute de le voir de mon vivant. Je ne suis pas pessimiste pour autant, je crois

dur comme fer que lorsque quelque chose est très difficile techniquement, mais pas impossible à cause d'une loi fondamentale de la physique, cela finit toujours par être réalisé. Car si jamais on butait sur une limite fondamentale, on aurait répondu à cette question essentielle du « seuil » entre le monde classique et le monde quantique. Pour le moment, nous n'en avons pas la moindre idée. Alors de deux choses l'une : soit on arrive à construire l'ordinateur quantique et ce sera merveilleux, soit on bute sur une limite, et tous les gens qui ont mis beaucoup d'argent dans ces recherches seront très déçus...⁸⁹

Si des obstacles scientifiques devaient remettre en question la possibilité de créer des *IA générales* telles que les promoteurs du transhumanisme les imaginent, cela pourrait constituer un coup d'arrêt aux « narratifs » qui ont jusqu'ici convaincu les investisseurs qu'un nouvel eldorado de l'IA était à portée de main. La conception par les architectes des technologies de ces *IA générales* reste encore lointaine. S'il est possible à des IA d'effectuer des manipulations complexes en prenant appui, lors de leur entraînement, sur de vastes ensembles de données, les tâches les plus simples qu'effectue un humain semblent encore hors de leur portée. Ainsi, en 1988, l'expert en robotique Hans Moravec évoquait les limites des IA en ces termes : « *Il est relativement facile de faire en sorte que les ordinateurs présentent des performances de niveau adulte lors de tests d'intelligence ou lorsqu'ils jouent aux dames, et difficile, voire impossible, de leur donner les compétences d'un enfant d'un an en matière de perception ou de mobilité.* »⁹⁰

Plus de trois décennies après ce constat, et malgré le développement à une vitesse inédite des grands modèles de langage, les limitations des IA demeurent un enjeu majeur de leur évolution. Selon Yan Le Cun, ancien directeur scientifique de l'IA chez *Meta* et lauréat du prix Turing : « *Nous n'avons pas de robot capable de*

⁸⁹ Alain Aspect, Prix Nobel de physique dans *Le Nouvel Esprit Public* (25 juin 2023) www.lenouvelespritpublic.fr/podcasts/446

⁹⁰ “It's comparatively easy to make computers exhibit adult-level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception or mobility.”

Moravec, Hans. *Mind children: The future of robot and human intelligence*. (Harvard University Press 1988) digitaleconomy.stanford.edu/news/the-turing-trap-the-promise-peril-of-human-like-artificial-intelligence/

*réaliser les mêmes tâches qu'un enfant de 5 ou 6 ans »*⁹¹. Cette constatation est la base de ses recherches visant à concevoir des nouvelles générations d'IA reposant sur des « *modèles du monde* » (ou *world models*). Ces IA seront fondées non seulement sur l'apprentissage du langage, mais aussi sur la modélisation du monde physique. Ce champ, désormais désigné sous le terme d'« *intelligence artificielle physique* », pourrait jouer un rôle déterminant dans le développement de robots intelligents. En effet, pour dépasser les applications industrielles actuelles, ces robots devront être capables de prendre en compte leur environnement et ainsi s'adapter pour exécuter des tâches complexes. Les « *IA physiques* » devraient aussi permettre de concevoir des systèmes capables de générer des simulations complexes et d'établir des prédictions basées sur l'analyse des données issues du monde réel (photos, vidéos, sons ou autres capteurs)⁹².

En effet, même les promoteurs des thèses transhumanistes comme Ray Kurzweil ou Nick Bostrom reconnaissent que les plus puissantes machines actuelles sont encore très éloignées des capacités du cerveau humain lorsqu'il est question d'interagir avec le monde réel. Pour la philosophe Susan Schneider, l'engouement face aux capacités nouvelles des IA ne doit pas masquer le fait qu'elles restent encore fondamentalement limitées :

*Bien sûr, la vitesse n'est pas tout. En dehors des calculs arithmétiques, votre cerveau est bien plus puissant que Summit [l'un des supercalculateurs d'IBM les plus rapides au monde]. Notre cerveau est le fruit de 3,8 milliards d'années d'évolution (l'âge estimé de la vie sur Terre) et a consacré sa puissance à la reconnaissance de formes, à l'apprentissage rapide et à d'autres défis pratiques de survie. Les neurones individuels sont peut-être lents, mais ils sont organisés de manière massivement parallèle, ce qui laisse encore les systèmes d'intelligence artificielle modernes dans la poussière*⁹³.

⁹¹ « Nous n'avons pas de robot capable de faire la même chose qu'un enfant de 5 ou 6 ans » : les « world models », nouvelle frontière de l'IA (Les Échos, 5 décembre 2025) www.lesechos.fr/tech-medias/intelligence-artificielle/nous-navons-pas-de-robot-capable-de-faire-la-meme-chose-quun-enfant-de-5-ou-6-ans-les-world-models-nouvelle-frontiere-de-lia-2202886

⁹² PAN: A World Model for General, Interactable, and Long-Horizon World Simulation (Institute of Foundation Models, Jiannan Xiang, 25 Nov 2025) arxiv.org/abs/2511.09057

⁹³ « Of course, speed is not everything. If the metric is not arithmetic calculations, your brain is far more computationally powerful than Summit. It is the product of 3.8 billion years of evolution (the estimated age of

Mais, au-delà même de la puissance de calcul ou des capacités d'adaptation face à des situations nouvelles, c'est la notion même qu'une machine puisse un jour acquérir une forme de conscience qui continue de faire débat au sein de la communauté des spécialistes de l'IA.

Ainsi, malgré les certitudes des promoteurs du transhumanisme et de la singularité technologique, des limitations fondamentales à la réalisation d'une *Intelligence Artificielle Générale* seront peut-être découvertes dans les années à venir. Et ce malgré les considérables moyens financiers, technologiques et humains qui sont consacrés au développement de ces IA. Mais avant même que ces IA puissent un jour être perçues ou non comme comparables à celles des humains, leurs effets sociaux, politiques et culturels seront déjà considérables sur nos sociétés.

B. De la science-fiction à la réalité industrielle

Si les IA actuelles sont encore très éloignées des créatures pensantes imaginées par les auteurs de science-fiction, qu'il s'agisse des intelligences artificielles de *2001: A Space Odyssey*, *Ex Machina* ou *HER*, pour nombre d'entre elles, les technologies que nous utilisons aujourd'hui ont d'abord été imaginées par des auteurs de science-fiction. Les créateurs de *Star Trek* sont ainsi à l'origine de nombreuses technologies que nous utilisons quotidiennement : les téléphones mobiles, les tablettes, les appareils de diagnostic mobiles, les interfaces vocales intelligentes ou encore les univers virtuels partagés. Le responsable de l'IA chez *Amazon* déclare même que l'ordinateur vocal de *Star Trek* représente la boussole pour la conception de leur IA vocale *Alexa*⁹⁴.

life on the planet) and has devoted its power to pattern recognition, rapid learning, and other practical challenges of survival. Individual neurons may be slow, but they are organized in a massively parallel fashion that still leaves modern AI systems in the dust. »

Susan Schneider : *Artificial You: AI and the Future of Your Mind* (Princeton University Press 2019)

⁹⁴ How Star Trek inspired Amazon's Alexa (VentureBeat, 7 Jun 2017)

venturebeat.com/ai/how-star-trek-inspired-amazons-alexa/

Les liens entre les industries de haute technologie et l'univers de la science-fiction n'ont sans doute jamais été aussi étroits qu'aujourd'hui. Mais, les auteurs de science-fiction imaginent aussi de nombreuses formes d'avenirs « dystopiques » : surveillance généralisée, déshumanisation, conditionnement des esprits ou encore manipulations génétiques de masse. Plutôt que l'émergence d'une IA malveillante, le défi des prochaines générations pourrait être d'éviter que des scénarios, à mi-chemin entre la dictature de surveillance de *Minority Report* et la société eugéniste de *Gattaca*, ne deviennent des réalités. Dans le cas de *Minority Report*, les oracles pourraient être interprétés comme un artifice narratif pour évoquer une version incarnée des prédictions de crimes établies par des modèles IA. Comme le remarquent Eric Schmidt, Henry Kissinger et Daniel Huttenlocher, lorsque les IA seront progressivement associées à l'ensemble de nos activités et de nos décisions, la soumission volontaire aux opinions ou prédictions données par ces IA ne relèvera plus de la science-fiction.

À mesure que l'intelligence artificielle intégrera la vie quotidienne, qu'elle l'étendra et la transformera, l'humanité sera confrontée à des pulsions contradictoires. Confrontés à des technologies qui dépassent la compréhension du profane, certains pourraient être tentés de considérer les déclarations de l'intelligence artificielle comme des jugements quasi divins. De telles pulsions, bien qu'erronées, ne manquent pas de sens. Dans un monde où une intelligence qui dépasse notre compréhension ou notre contrôle tire des conclusions utiles mais étrangères, est-il insensé de s'en remettre à ses jugements ? Stimulé par cette logique, un réenchantement du monde pourrait s'ensuivre, dans lequel les IA seraient considérées comme des oracles auxquels certains humains s'en remettraient sans poser de questions⁹⁵.

⁹⁵ As AI is woven into the fabric of daily existence, expands that existence, and transforms it, humanity will have conflicting impulses. Confronted with technologies beyond the comprehension of the nonexpert, some may be tempted to treat AI's pronouncements as quasi-divine judgments. Such impulses, though misguided, do not lack sense. In a world where an intelligence beyond one's comprehension or control draws conclusions that are useful but alien, is it foolish to defer to its judgments? Spurred by this logic, a re-enchantment of the world may ensue, in which AIs are relied upon for oracular pronouncements to which some humans defer without question.

Henry A. Kissinger, Eric Schmidt and Daniel Huttenlocher. « The Age of AI: And Our Human Future. » (John Murray Press, 2021)

Cependant, il reste des domaines politiques ou de haute sécurité dans lesquels cette soumission volontaire aux avis de l'IA constituerait un risque majeur. Ainsi, Henri Kissinger, qui a longuement participé aux négociations internationales sur le désarmement, rappelle que les tractations entre blocs (en particulier les États-Unis et l'Union soviétique) ont eu lieu dès que les puissances dotées de l'arme nucléaire se sont affrontées. Nous sommes encore loin de pouvoir assister à de telles négociations pour la limitation du rôle offensif des IA. Et cela d'autant plus qu'elles peuvent être utilisées par des organisations ou des groupes agissant pour les États, mais en gardant une capacité de déni plausible (*plausible deniability*) pour les puissances qui les auront mandatés. Les problèmes liés à l'introduction d'IA dans la chaîne de commandement nucléaire rendent des négociations probables sur ce segment spécifique et hautement controversé de l'usage des IA. Ben Buchanan, ancien conseiller pour l'IA à la Maison-Blanche, confirme que, pour l'instant, les technologies d'IA ne sont pas en mesure d'être adaptées à des théâtres de conflits lorsqu'il est question des systèmes d'armes nucléaires.

Étant donné que les applications militaires de l'IA privilégiées par les militaires sont distinctes des applications scientifiques privilégiées par les évangélistes, la plupart des systèmes d'apprentissage automatique ne sont pas prêts pour des environnements hostiles, comme la guerre contre des adversaires dotés de bonnes capacités de piratage. Des tests et des évaluations beaucoup plus itératifs doivent être effectués avant que ces systèmes puissent être considérés comme éthiquement et stratégiquement appropriés au combat. En raison de ces préoccupations majeures concernant la sécurité et la robustesse, les utilisations de l'apprentissage automatique pour des enjeux majeurs – comme dans les systèmes de commandement et de contrôle nucléaires – sont manifestement imprudentes dans un avenir prévisible.⁹⁶

⁹⁶ Since the military applications of AI favored by the warriors are distinct from the scientific applications preferred by the evangelists, most machine learning systems are not ready for adversarial environments like war against opponents with good hacking capabilities; much more iterative testing and evaluation must be done before these systems can be considered ethically and strategically appropriate for combat.⁴⁶ Due to these major concerns about security and robustness, some high-stakes uses of machine learning—such as in nuclear command-and-control systems—are manifestly unwise for the foreseeable future.

Buchanan, Ben; Imbrie, Andrew. *The New Fire: War, Peace, and Democracy in the Age of AI* (p. 244). MIT Press 2022

Les experts en sécurité nucléaire du *European Leadership Network* considèrent également que les risques seraient trop élevés pour envisager l'utilisation de systèmes d'IA pour le déclenchement du feu nucléaire. Ils recommandent même qu'un moratoire international soit mis en place dans ce domaine :

Du point de vue de la sécurité mondiale, l'IA présente des risques d'erreurs de calcul, de perception et d'interprétation erronées, qu'elles soient dues à des défaillances du système, à des vulnérabilités ou à une mauvaise utilisation, pouvant conduire à une escalade involontaire ou accidentelle. Si ces modèles sont intégrés dans des systèmes de prise de décision stratégique, ils présentent certains des risques les plus graves en raison de leur nature opaque, de leur imprévisibilité et de leur vulnérabilité aux cyberattaques. La nature « boîte noire » de certains modèles d'IA rend leur processus de prise de décision très difficile à déchiffrer avec notre compréhension actuelle. S'ils sont intégrés dans des systèmes de prise de décision stratégique, cela pourrait ne laisser aucun système de responsabilisation et aucune méthode de vérification des prédictions et des décisions de l'IA. Compte tenu des risques importants posés par les modèles de pointe en matière d'apprentissage profond, les États du P5 [France, Chine, États-Unis, Royaume-Uni, Russie] devraient imposer un moratoire sur l'intégration de ces modèles dans des domaines spécifiques du NC3 [Nuclear Command, Control, and Communications] présentant les risques d'intégration les plus élevés⁹⁷.

⁹⁷ From a global security perspective, AI introduces risks of miscalculations, misperceptions and misinterpretations, whether from system failures, vulnerabilities, or misuse, potentially leading to inadvertent or accidental escalation [...] If incorporated into strategic decision-making systems, these models pose some of the gravest risks due to their opaque nature, unpredictability, and susceptibility to cyber-attacks[...] The 'black box' nature of some AI models makes its decision-making process very challenging to decipher with our current understanding of these models [...] If integrated in strategic decision-making systems, this might leave no accountability systems and no method of verification for AI predictions and decisions. Considering the significant risks posed by cutting-edge deep learning models, P5 states should impose a moratorium on the integration of these models into specific NC3 areas with the highest integration risks.

Alice Saltini : AI and nuclear command, control and communications: P5 perspectives (The European Leadership Network Nov 2023) www.europeanleadershipnetwork.org/wp-content/uploads/2023/11/AVC-Final-Report-online-version.pdf

C. IA, données et modèles économiques toxiques

Les modèles économiques basés sur l'attention et sur l'hyper-individualisation des messages ont pour corollaire l'extraction toujours plus grande d'informations sur les individus. Il existe désormais une tension entre les objectifs des industriels de l'IA qui souhaitent utiliser des jeux de données de plus en plus larges pour améliorer leurs modèles d'IA et les citoyens qui pourraient être dépossédés de leurs données personnelles. Ils ne seraient plus en mesure de contrôler ce qu'il adviendra de ces données (et donc d'eux-mêmes) par la suite. Comme le décrit Tomas Chamorro-Premuzic, spécialiste de la psychologie des organisations, un cercle vicieux issu du modèle économique des plateformes publicitaires s'est ainsi mis en place. Cette évolution a eu des conséquences à la fois individuelles, sociales et politiques :

La quantification de notre attention – et la capacité de l'IA à utiliser cette information comme une arme – crée un cercle vicieux : notre attention étant rare et les informations abondantes, la bataille pour notre attention s'est accélérée. Netflix est en concurrence avec Twitter, Twitter est en concurrence avec le New York Times, et le New York Times est en concurrence avec Instagram ; tous sont en concurrence pour notre précieux temps et notre encore plus précieuse attention. Les algorithmes de ces plateformes ont soif de notre attention, et leurs modèles économiques en dépendent, ce qui rend notre attention très précieuse, notamment parce qu'il en reste très peu à capter une fois que ces algorithmes l'ont consommée. Cela conduit également à une surcharge d'informations encore plus importante, qui menace de nous distraire encore davantage. Le résultat est une dégradation de la concentration qui provoque des troubles déficitaires de l'attention avec ou sans hyperactivité (TDAH), de l'hyperactivité agitée, de l'ennui rapide ou encore de l'impulsivité⁹⁸.

⁹⁸ The quantification of our attention—and AI's ability to weaponize that information—creates a vicious cycle: since our attention is scarce and information is plentiful, the battle for our attention exacerbates. Netflix is competing with Twitter, Twitter is competing with the New York Times, and the New York Times is competing with Instagram; they're all competing for our precious time and our even more precious focus. Their algorithms crave our attention, and their business models depend on it, which makes our attention highly valuable, not least because there is so little of it left to capture after the algorithms consume it. It also leads to yet more information overload, which threatens to distract us even more. The result is a degradation of focus that causes

On remarque ainsi une lacune dans le raisonnement de nombreux spécialistes de l'IA comme Neil Lawrence⁹⁹, qui ne peuvent (ou ne souhaitent pas) imaginer que des restrictions sur l'accès aux données personnelles soient mises en place. Seules des formes de compensation minimales sont envisagées de la part des sociétés qui utiliseraient ces données à leur profit, ce qui a pour conséquence de prolonger une analogie tentante mais erronée : les données sont perçues comme le « nouveau pétrole » de l'économie du XXI^e siècle. À l'inverse, Shoshana Zuboff souhaite imposer des restrictions sur le recueil massif de ces données et donc sur le modèle économique de ces plateformes. Le recueil toujours plus massif de données sur les individus constitue désormais un risque politique de manipulation mais aussi un risque économique, au point que même les experts du *Fonds monétaire international* s'en inquiètent désormais : « *La plupart des transactions impliquant des données personnelles se font à l'insu des utilisateurs, qui n'ont probablement même pas conscience qu'elles ont eu lieu et, a fortiori, qu'ils ont donné leur autorisation. Cela engendre ce que les économistes appellent une externalité : le coût de la perte de vie privée n'est pas parfaitement pris en compte lorsqu'un échange de données a lieu. La conséquence est que l'opacité du marché conduit sans doute à collecter trop de données et à partager une trop faible fraction de leur valeur avec les individus*¹⁰⁰. »

Comme l'ont montré les scandales liés à l'utilisation incontrôlée des données par les grandes plateformes de l'Internet, ces sociétés ne peuvent répondre spontanément à des principes démocratiques lorsque ceux-ci s'inscrivent en contradiction directe avec leurs objectifs industriels et financiers. Un exemple de ces conduites d'entreprises, qui, comme *Facebook*, ont favorisé la radicalisation

attention deficit hyperactivity disorder (ADHD)–like behaviors, such as restless hyperactivity, rapid boredom, and impulsivity

Chamorro-Premuzic, Tomas. I, Human: AI, Automation, and the Quest to Reclaim What Makes Us Unique (p. 40) (Harvard Business Review Press 2023)

⁹⁹ Neil D. Lawrence. *The Atomic Human: What Makes Us Unique in the Age of AI* (Public Affairs 2024)

¹⁰⁰ Bâtissons ensemble une meilleure économie des données par Yan Carrière-Swallow et Vikram Haksar (Fonds Monétaire International Mars 2021)

www.imf.org/external/pubs/ft/fandd/fre/2021/03/pdf/how-to-build-a-better-data-economy-carriere.pdf

idéologique de leurs usagers, a été donné par Max Fisher dans son ouvrage *The Chaos Machine* :

J'ai appris que, peu de temps avant ma visite, certains chercheurs de Facebook, qui avaient été désignés en interne pour étudier les effets de leur technologie en réponse aux soupçons croissants selon lesquels le site pourrait aggraver les divisions politiques de l'Amérique, avaient averti la société que la plateforme faisait exactement ce que ses dirigeants avaient, lors de nos conversations, ignoré d'un revers de la main. « Nos algorithmes exploitent l'attraction du cerveau humain pour la discorde », ont averti les chercheurs dans une présentation de 2018 qui a été divulguée au Wall Street Journal. En fait, poursuivait la présentation, les systèmes de Facebook étaient conçus de manière à fournir aux utilisateurs « un contenu de plus en plus conflictuel dans le but d'attirer l'attention des utilisateurs et d'augmenter le temps passé sur la plateforme ». Les dirigeants ont mis la recherche de côté et ont largement rejeté ses recommandations, qui préconisaient de modifier les systèmes promotionnels qui choisissent ce que les utilisateurs voient et qui auraient pu réduire leur temps passé en ligne¹⁰¹.

Ces préoccupations morales, lorsqu'elles ne sont pas intégrées en amont par les concepteurs d'IA, et surtout si elles sont délibérément écartées, peuvent avoir des conséquences durables sur les usagers de ces IA. Cette question est ainsi résumée par Steven Kerr, de l'Université de l'Ohio : « *Si le système de récompense est conçu de telle manière qu'il est irrationnel d'être moral, cela ne signifie pas nécessairement que l'immoralité en résultera. Mais n'est-ce pas là chercher les ennuis ?*¹⁰² »

¹⁰¹ I later learned that, a short time before my visit, some Facebook researchers, appointed internally to study their technology's effects, in response to growing suspicion that the site might be worsening America's political divisions, had warned internally that the platform was doing exactly what the company's executives had, in our conversations, shrugged off. "Our algorithms exploit the human brain's attraction to divisiveness," the researchers warned in a 2018 presentation later leaked to the Wall Street Journal. In fact, the presentation continued, Facebook's systems were designed in a way that delivered users "more and more divisive content in an effort to gain user attention & increase time on the platform." Executives shelved the research and largely rejected its recommendations, which called for tweaking the promotional systems that choose what users see in ways that might have reduced their time online.

Max Fisher. *The Chaos Machine The Inside Story of How Social Media Rewired Our Minds and Our World* (pp. 14-15). (Quercus Editions Ltd 2022)

¹⁰² If the reward system is so designed that it is irrational to be moral, this does not necessarily mean that immorality will result. But is this not asking for trouble? —Steven Kerr cité par Brian Christian.

VIII. IA ET SANTÉ : STRATÉGIES ET ENJEUX POLITIQUES

A. Données de santé : perspectives industrielles et politiques

L'évolution des technologies d'IA, en plus d'aider au diagnostic des pathologies, rend également possible l'analyse des données issues des capteurs présents dans notre environnement à des fins de détection précoce des pathologies. Cette démarche peut exister à l'échelle des individus, mais aussi à l'échelle de populations entières. Ainsi, avant même la pandémie de *Covid-19*, de nombreux scientifiques avaient évoqué la possibilité d'intégrer des réseaux de capteurs associés à des systèmes d'IA pour effectuer la détection précoce des épidémies et ainsi développer des mesures de prévention. À l'extrême, pour le cabinet d'étude *Frost & Sullivan*, la réponse la plus « efficace » en termes de lutte contre les pandémies consisterait à établir un réseau mondial de capteurs. Ce réseau permettrait de détecter de manière précoce les menaces biologiques où qu'elles apparaissent dans le monde. Cependant, comme le notent les auteurs, cette proposition, qui correspondrait à la plus importante opportunité de marché technologique jamais conçue pour l'Internet des Objets (*IdO* ou *IoT* en anglais), se heurterait à l'opposition des opinions publiques en raison de son caractère liberticide :

La solution la plus simple serait de permettre aux entreprises, aux villes et aux gouvernements de créer collectivement un vaste réseau mondial de capteurs pour détecter les virus. Cependant, cela nécessiterait une planification et une mise en œuvre à l'échelle mondiale qui s'attaquerait aux fondements mêmes de la démocratie et exigerait des gouvernements qu'ils placent les besoins de la

The Academy of Management Executive (1993-2005) Vol. 9, No. 1 pp. 7-14 (Feb 1995)
web.mit.edu/curhan/www/docs/Articles/15341_Readings/Motivation/Kerr_Folly_of_rewarding_A_while_hoping_for_B.pdf

planète devant ceux de leurs citoyens. La solution la plus logique est souvent la plus difficile à mettre en œuvre. Le degré de planification nécessaire pour réaliser cette solution en ferait l'une des réalisations les plus importantes de l'histoire de l'humanité [...] et, à long terme, cela représente le « Saint Graal » des opportunités de marché pour l'IoT¹⁰³.

La fascination pour la « techno-efficience » de certains acteurs technologiques rappelle le « *solutionnisme technologique* » évoqué par Evgeny Morozov dans son ouvrage « *Pour tout résoudre cliquez ici* »¹⁰⁴. Cette fascination est encore plus présente lorsqu'il est question des technologies de l'IA. Cette tendance n'est pas uniquement le fait de régimes autoritaires, mais bien d'acteurs économiques ou de gouvernements qui évaluent le rapport bénéfice/risque de ces technologies et considèrent que démocratie et protection des libertés peuvent devenir des variables d'ajustement.

À mesure que se développent de nouvelles méthodes d'analyse basées sur l'intelligence artificielle, les technologies de l'Internet des objets ont pu se diversifier. Qu'il s'agisse d'analyser les données issues d'objets connectés portables (ou *wearables* : casques, bagues, lunettes ou bracelets dotés de capteurs), de drones pour l'imagerie thermique ou encore de réseaux de capteurs urbains¹⁰⁵.

¹⁰³ The simple answer might for enterprises, cities, and national governments to collectively create a massive global network of sensors to detect viruses. However, this would require planning and implementation on a global scale that would tax the very foundations of democracy and obligate governments to place the needs of the planet ahead of the needs of their citizens. The most logical solution is often the most difficult to implement. The amount of planning required to make this solution a reality would, arguably, make it one of the most significant achievements in the history of mankind. [...] I find this to be the “holy grail” of IoT opportunity in the long term.

The Next Generation of IoT – Addressing the Coronavirus and Preventing Future Outbreaks (Frost & Sullivan, Jan 31, 2020) www2.frost.com/frost-perspectives/the-next-generation-of-iot-addressing-the-coronavirus-and-preventing-future-outbreaks/

¹⁰⁴ Pour tout résoudre cliquez ici - l'aberration du solutionnisme technologique (Evgeny Morozov, Ed. Fyp 2014). Ouvrage original : To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist (Evgeny Morozov, Ed. Penguin 2013)

¹⁰⁵ Internet of Things for Current COVID-19 and Future Pandemics : An Exploratory Study by M. Nasajpour, S. Pouriyeh, M. Parizi, M. Dorodchi, M. Valero, H. Arabnia Dept of Information Technology and Dept of Software Engineering and Game Development, Kennesaw State University, Department of Computer Science, University of North Carolina and University of Georgia (submitted 22 Jul 2020) <https://arxiv.org/pdf/2007.11147>

Les IA peuvent ainsi révéler (ou déduire) des éléments sur l'état de santé d'une personne à partir d'informations en apparence anodines. Ainsi, lorsque l'on analyse dans la durée les déplacements d'une personne, on peut anticiper la survenue possible de troubles cardiaques ou circulatoires. *Facebook* a ainsi développé des brevets portant sur l'analyse en continu des déplacements et des comportements qui permettent d'évaluer les paramètres de santé des

“ *La seule intimité qui vous reste est ce qui se passe à l'intérieur de votre crâne. Et peut-être est-ce suffisant...*

Extrait de dialogue du
film "*Enemy of the State*"
Tony Scott 1998

utilisateurs du réseau social¹⁰⁶. En 1998, dans le film *Enemy of the State* de Tony Scott, le directeur adjoint de la NSA, joué par Jon Voight, affirmait qu'en matière de surveillance : « *La seule intimité qui vous reste est ce qui se passe à l'intérieur de votre crâne. Et peut-être est-ce suffisant...*¹⁰⁷ ». Cependant, cet obstacle ultime à la connaissance des pensées des individus pourrait lui-même être surmonté dans les années à venir. Des technologies d'IA pourraient, en effet être en mesure d'analyser la formation des images mentales directement dans le

cerveau. À terme, ces innovations, qui suscitent l'intérêt des chercheurs lorsqu'il est question de créer de nouveaux canaux de communication avec des personnes handicapées, pourraient connaître des applications « non médicales » pour analyser les processus cérébraux. Il deviendrait alors possible d'influencer, de manière imperceptible, le comportement ou les pensées des individus. Ainsi, en Chine, plusieurs entreprises expérimentent des casques qui analysent les ondes cérébrales de leurs employés afin de détecter le stress, la colère ou

¹⁰⁶ What 7 Creepy Patents Reveal about Facebook (New York Times, 21 Jun 2018)
[nytimes.com/interactive/2018/06/21/opinion/sunday/facebook-patents-privacy.html](https://www.nytimes.com/interactive/2018/06/21/opinion/sunday/facebook-patents-privacy.html)

¹⁰⁷ "The only privacy that's left is the inside of your head. Maybe that's enough."

Enemy of the State (Movie Transcript)

<http://movietranscript.blogspot.com/2016/01/1998-enemy-of-state-english-transcripts.html>

l'endormissement¹⁰⁸. Les laboratoires du groupe *Meta* commencent également à utiliser des modèles d'IA générative pour recréer les images perçues par des personnes en analysant les modifications cérébrales captées par des systèmes d'imagerie par résonance magnétique (*IRM*)¹⁰⁹. De même, des capteurs électroencéphalographiques associés à des systèmes d'IA sont désormais utilisés pour analyser l'activité cérébrale afin de découvrir les mots formés dans le cerveau¹¹⁰.

Là encore, les mesures prises pour assurer la sécurité des données recueillies par ces capteurs détermineront si ces technologies s'inscrivent dans une trajectoire de prévention, de contrôle ou de manipulation des individus. Comme le notent plusieurs auteurs, la limite entre prévention des conduites à risque et conditionnement des utilisateurs via les objets connectés reste délicate à fixer. Cette tendance au contrôle des populations pourrait en effet marquer le passage de l'ingénierie sociale au contrôle social assisté par l'IA à l'instar du *Crédit social* mis en place en Chine¹¹¹.

¹⁰⁸ En Chine, des capteurs cérébraux pour surveiller les émotions des employés (Slate 1 mai 2018) www.slate.fr/story/161173/en-chine-des-capteurs-cerebraux-pour-surveiller-les-emotions-des-employes

¹⁰⁹ Meta recreates mental imagery from brain scans using AI (Venture Beat, 18 Oct 2023) venturebeat.com/ai/meta-recreates-mental-imagery-from-brain-scans-using-ai/

¹¹⁰ Scientists use AI to decipher words and sentences from brain scans (Science.org, 1 May 2023) www.science.org/content/article/scientists-use-ai-decipher-words-and-sentences-brain-scans

¹¹¹ The Social Credit System: Not Just Another Chinese Idiosyncrasy (Eunsun Cho - Journal of Public and International Affairs, Princeton University, 1 May 2020) <https://jpia.princeton.edu/news/social-credit-system-not-just-another-chinese-idiosyncrasy>

B. Modifier le centre de gravité de l'économie de la santé

Avec la montée en puissance des technologies d'IA, la captation et le traitement des données de santé sont devenus des enjeux stratégiques pour les acteurs des technologies. Les dispositifs de santé connectée, et en particulier les technologies de suivi et de prévention des pathologies, deviennent le nouveau terrain d'expansion pour les industriels des technologies. Grâce à de nouvelles générations d'objets médicaux connectés capables de suivre en continu les activités et les paramètres physiologiques des usagers, il devient possible de mettre en place à moindre coût des systèmes de prévention et de suivi personnalisés. Ces objets connectés qui accompagneront les usagers au quotidien devraient permettre le diagnostic précoce de pathologies chroniques (cancer, diabète, asthme, maladies cardiovasculaires...).

Ainsi, après avoir intégré des fonctions d'électrocardiogramme, les récentes versions de l'*Apple Watch* devraient, grâce à des systèmes d'IA, détecter les apnées du sommeil via l'analyse des sons et des mouvements de leurs utilisateurs. Ces objets connectés de santé devraient également permettre de modifier les comportements de leurs utilisateurs pour maîtriser les facteurs de risque, comme la sédentarité ou l'obésité.

Ainsi, le secteur de l'assurance santé est devenu l'une des cibles prioritaires des grands acteurs des technologies. Ces acteurs technologiques peuvent ainsi prendre pied dans le secteur des assurances et déplacer le centre de gravité de l'économie de la santé vers la prévention. En effet, les dépenses de santé sont actuellement massivement axées sur les soins¹¹². Grâce aux évolutions des nouvelles générations de services d'IA liées à la santé connectée, il deviendrait possible pour ces sociétés de générer un marché de taille mondiale pour la prévention en santé.

¹¹² Santé : « En France, la culture de la prévention n'est clairement pas acquise » (David Simard, Le Monde 27 août 2023) www.lemonde.fr/idees/article/2023/01/05/sante-en-france-la-culture-de-la-prevention-n-est-clairement-pas-acquise_6156771_3232.html

Voir, sur ce point, le rapport de *Goldman Sachs*¹¹³ sur la santé connectée qui estimait à 305 milliards de dollars les économies qui seraient induites aux États-Unis par l'introduction des objets connectés dans le domaine de la santé. Parmi ces économies, 200 milliards de dollars seraient liés à l'amélioration de la prévention et de la gestion des pathologies chroniques, en particulier les maladies cardiovasculaires, l'asthme et le diabète. Ces économies représenteraient aux États-Unis près de 10 % du total des dépenses de santé (soit 4300 milliards de dollars en 2023)¹¹⁴.

La tentation d'étendre les mesures de prévention jusqu'à les rendre potentiellement liberticides est d'autant plus grande que ces mesures pourraient avoir un impact significatif sur l'équilibre économique des pays développés. Ainsi, cette dynamique de prévention des risques était déjà présente dans la proposition de loi *HR1313* introduite en 2017 au *Congrès américain* par l'administration Trump. L'objectif affiché de cette proposition de loi était, grâce à des tests génétiques réalisés à grande échelle en entreprise, de développer des mesures de prévention et de détection précoce des maladies. Cette loi prévoyait de déployer ces tests génétiques dans les entreprises américaines et d'imposer des pénalités de 5 000 dollars par an aux employés qui refuseraient de se soumettre à ce « screening génétique »¹¹⁵. L'opposition des élus démocrates au *Congrès américain* n'a cependant pas permis à ce projet d'être adopté.

¹¹³The Digital Revolution comes to US Healthcare (Goldman Sachs Equity Research 2015)
www.anderson.ucla.edu/Documents/areas/adm/acis/library/DigitalRevolutionGS.pdf

¹¹⁴ National Health Expenditures 2023 (U.S. Centers for Medicare & Medicaid Services 2019)
www.cms.gov/data-research/statistics-trends-and-reports/national-health-expenditure-data/nhe-fact-sheet

¹¹⁵ Employees who decline genetic testing could face penalties under proposed bill (Washington Post, 11 mars 2017) www.washingtonpost.com/news/to-your-health/wp/2017/03/11/employees-who-decline-genetic-testing-could-face-penalties-under-proposed-bill/

C. Quelle stratégie pour les données de santé en Europe ?

Dans le domaine de l'IA appliquée à la santé, la création par le ministère de la Santé d'une plateforme de données de santé (*Health Data Hub*) constitue un volet important de la stratégie française en matière d'ouverture des données de santé. Cette plateforme a en effet été conçue pour devenir un guichet unique d'accès à l'ensemble des données de santé des Français. Elle vise à favoriser le développement de nouveaux services d'IA appliquée à la santé qu'il s'agisse de l'aide au diagnostic, du suivi des traitements ou du développement de mesures de prévention. Cependant, comme l'ont fait remarquer des professionnels de santé ainsi que des spécialistes de l'informatique médicale, l'hébergement de ce dispositif par la société *Microsoft*¹¹⁶ constitue à la fois un risque en termes de souveraineté sur des données sensibles et une opportunité manquée pour développer des savoir-faire essentiels dans l'écosystème français de la santé numérique¹¹⁷. La *CNIL* a même estimé en 2020 que « *le changement de la solution d'hébergement du Health Data Hub et des autres entrepôts de santé hébergés par les sociétés soumises au droit étasunien devrait intervenir dans un délai aussi bref que possible*¹¹⁸ ». L'accès aux données de santé constitue un enjeu d'autant plus stratégique pour les acteurs des technologies que les données de santé des Français figurent parmi les collections les plus denses et les plus structurées de l'ensemble des pays développés. Ainsi, en raison de leur qualité et de leur étendue en France et en Europe, ces données de santé constituent un levier puissant pour entraîner les modèles d'IA et les aider à devenir plus performants.

¹¹⁶ « L'exploitation de données de santé sur une plate-forme de Microsoft expose à des risques multiples » (tribune parue dans Le Monde du 10 décembre 2019) www.lemonde.fr/idees/article/2019/12/10/l-exploitation-de-donnees-de-sante-sur-une-plate-forme-de-microsoft-expose-a-des-risques-multiples_6022274_3232.html

¹¹⁷ Health Data Hub : « Le choix de Microsoft, un contresens industriel ! » (Interview de Bernard Benhamou Le Point, 18 Juin 2020) www.lepoint.fr/technologie/health-data-hub-le-choix-de-microsoft-et-un-contresens-industriel-10-06-2020-2379394_58.php

¹¹⁸ La Cnil réclame l'arrêt de l'hébergement des données de santé des Français par Microsoft (L'Obs, 9 octobre 2020) www.nouvelobs.com/high-tech/20201009.OBS34527/la-cnil-reclame-l-arret-de-l-hebergement-des-donnees-de-sante-des-francais-par-microsoft.html

Face aux controverses liées aux choix de la société *Microsoft*, le gouvernement a, dans un premier temps, annoncé qu'il comptait rapatrier l'hébergement vers des sociétés françaises ou européennes¹¹⁹. Cependant, malgré des démarches de parlementaires visant à obtenir des informations sur l'état du projet de migration de cette plateforme¹²⁰, cette migration n'a toujours pas eu lieu. La protection des données de santé pourrait avoir une dimension encore plus cruciale lors de la mise en place de *l'Espace européen des données de santé*¹²¹ dans le cadre de la stratégie européenne pour les données¹²². À ce titre, la récente décision de la *CNIL* d'autoriser *Microsoft* à héberger le projet européen *EMC2* constitue un précédent préoccupant, au vu des risques sur la confidentialité des données de santé que font courir les lois extraterritoriales américaines, et en premier lieu la loi *FISA (Foreign Intelligence Surveillance Act)*¹²³. Grâce à cette loi, les agences de renseignement américaines peuvent requérir les données traitées par des sociétés américaines, où que ces données soient situées dans le monde¹²⁴.

L'un des enjeux stratégiques pour la puissance publique en France et en Europe sera de veiller à ce que les évolutions de l'IA en santé ne remettent pas en cause notre modèle social au profit d'une logique de contrôle systématisé des individus. Ce risque sera d'autant plus grand que seront bientôt intégrées à ces plateformes des données issues de la génomique. La mise en place d'une logique d'hyper-individualisation de la couverture santé irait à l'encontre de notre modèle social fondé sur la solidarité et la mutualisation des risques au niveau de la société tout

¹¹⁹ Microsoft doit se retirer du Health Data Hub, d'après la Cnil (L'Usine Digitale 9 octobre 2020)

www.usine-digitale.fr/article/microsoft-doit-se-retirer-du-health-data-hub-d-apres-la-cnil.N1014634

¹²⁰ Le député Philippe Latombe saisit la CADA pour obtenir le "benchmark" de l'hébergement du Health Data Hub (TIC Pharma 11 mai 2023) www.ticpharma.com/story?ID=2268

¹²¹ European Health Data Space health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

¹²² Une stratégie européenne pour les données (Communication de la Commission au Parlement Européen, au Conseil, au Comité Économique et Social Européen et au Comité des Régions du 19 février 2020. <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52020DC00666>

¹²³ HDH: des fournisseurs français attaquent la décision de la Cnil favorisant Microsoft (Le Monde Informatique 18 mars 2024) www.lemondeinformatique.fr/actualites/lire-hdh-des-fournisseurs-francais-attaquent-la-decision-de-la-cnil-favorisant-microsoft-93256.html

¹²⁴ L'extension des prestataires américains devant collaborer avec la NSA fait polémique (Jean-Marc Manach - Next 22 avril 2024) <https://next.ink/135019/extension-des-prestataires-americains-devant-collaborer-avec-la-nsa-fait-polemique/>

entière. Désormais, la protection de notre modèle de protection sociale relève aussi de la souveraineté numérique. Or, les réseaux sociaux associés aux objets connectés dédiés à la santé permettent déjà aux grandes plateformes de compléter leurs informations sur les utilisateurs pour affiner leurs profils. Cela permet, comme jamais auparavant, de développer des modèles d'IA dédiés à l'évaluation précise du risque santé pour chaque individu. Ainsi, comme le précise Pascal Demurger, directeur général de la MAIF : « *C'est un renversement complet du monde de l'assurance. Traditionnellement, les assureurs avaient très peu de données sur leurs clients mais un grand nombre de clients. Grâce au big data, nous pouvons désormais récolter un grand nombre de données comportementales sur une seule personne* ¹²⁵ ».

Plutôt que de devenir directement des acteurs du soin, secteur nécessitant des investissements importants et risqués sur le long terme, les grandes plateformes de l'Internet peuvent désormais utiliser les données accumulées sur leurs usagers pour fournir des services d'assurance et de prévention des pathologies. Grâce aux données qu'elles détiennent sur leurs usagers, ces plateformes peuvent modéliser de manière précise les risques liés à la santé de chaque individu et ainsi optimiser les profits de leurs services d'assurances. *Google* (via *Alphabet*) s'est déjà lancé dans le secteur de l'assurance santé avec sa division *Coefficient* ¹²⁶. Les services liés à la santé et au bien-être constituent l'un des axes stratégiques d'*Apple* pour leurs montres connectées ainsi que leurs smartphones. Par ailleurs, *Facebook* développe des recherches sur les applications médicales des univers virtuels dans le cadre de son projet de *métavers* ¹²⁷.

¹²⁵ Santé: faut-il faire payer les assurés en fonction de leur mode de vie ? (Le Monde, 6 septembre 2016) www.lemonde.fr/economie/article/2016/09/06/assurance-votre-vie-privee-vaut-bien-une-ristourne_4993378_3234.html

¹²⁶ Verily (Alphabet) se lance dans l'assurance avec sa nouvelle division, Coefficient (L'Usine Digitale, 25 août 2020) www.usine-digitale.fr/article/verily-alphabet-se-lance-dans-l-assurance-avec-sa-nouvelle-division-coefficient.N996864

¹²⁷ Meta is forcing Apple into virtual reality (Quartz 11 Mar 2022) <https://qz.com/2141093/meta-is-using-quest-2-to-force-apple-into-the-metaverse-early>

Les technologies d'IA pourraient exercer un effet de levier sur l'ensemble des activités de nos systèmes de santé. Si des plateformes extra-européennes conçoivent l'ensemble des technologies d'IA liées au soin, à la prévention et à l'assurance santé en Europe, elles pourraient remettre en cause nos modèles sociaux au profit de démarches basées uniquement sur la rationalisation économique. Cela pourrait aboutir à la mise en place de contraintes en matière de prévention *via* des variations du prix des assurances santé en fonction du comportement des individus. En l'absence d'une politique industrielle qui seule permettrait le développement d'alternatives européennes, ces évolutions du système de soin pourraient s'accompagner d'effets collatéraux sous la forme d'atteintes aux droits des personnes et de dérives liées à la surveillance.

IX. ARCHITECTURE ET GOUVERNANCE DE L'IA

Les grands modèles de langage (*LLM*) commerciaux ont, dans un premier temps, été développés de manière verticale et centralisée. Chaque génération de modèle intègre davantage de paramètres (plusieurs milliards, voire milliers de milliards de paramètres pour les modèles les plus récents). À titre d'exemple :

- En mai 2020, *GPT-3* comptait 175 milliards de paramètres et a été entraîné sur 45 téraoctets de données textuelles.

- En mars 2023, *GPT-4* comptait plus de 1 000 milliards de paramètres et a été entraîné sur plus de 1 000 téraoctets de données.¹²⁸

Cette course à la puissance a été marquée par la compétition entre les principaux acteurs du cloud qui ont mis à contribution leurs puissances de calcul et de stockage pour développer de nouvelles générations d'IA, ce qui a fait craindre la constitution d'un duopole entre *Microsoft* et *Google* autour de l'IA¹²⁹. Comme le rappelle Tim Wu, le spécialiste de la régulation des entreprises technologiques américaines, dans son ouvrage *La malédiction de la grandeur (The Curse of Bigness)*¹³⁰, les risques liés aux monopoles ne sont pas uniquement économiques mais bien politiques. Les principaux acteurs des technologies exercent déjà un pouvoir considérable sur l'ensemble des activités humaines et sur le fonctionnement des démocraties elles-mêmes. Avec le développement des IA, et la concentration extrême des données autour de ces grandes plateformes, leur pouvoir pourrait être encore plus grand.

Cependant, la constitution de ces nouvelles « cathédrales » du code et des données des IA pourrait se heurter à de nouveaux obstacles économiques et politiques. Les ressources financières, humaines et énergétiques nécessaires à la constitution de

¹²⁸ LLMs hitting 2 trillions parameters (IEEE Future Directions 14 Nov 2023)

<https://cmt.ee.org/futuredirections/2023/11/14/llms-hitting-2-trillions-parameters/>

¹²⁹ The AI Boom That Could Make Google and Microsoft Even More Powerful (Wall Street Journal, 11 Feb 2023) www.wsj.com/articles/the-ai-boom-that-could-make-google-and-microsoft-even-more-powerful-9c5dd2a6

¹³⁰ Tim Wu. *The Curse of Bigness: Antitrust in the New Gilded Age* (Columbia Global Reports 2018)

ces IA pourraient rendre leur équilibre économique plus difficile à atteindre. Qu'il s'agisse des coûts liés à l'acquisition des processeurs dédiés (*GPU*), de l'entraînement et du fonctionnement de ces modèles dans des data centers ou de l'énergie nécessaire à leur fonctionnement.

Si les investisseurs ont, dans un premier temps, été stimulés par la perspective de la création d'une « *IA générale* », ils commencent à émettre des doutes quant à la réalisation des objectifs technologiques et économiques des grands modèles de langage (*LLM*)¹³¹. Les entreprises qui conçoivent les plus grands modèles de langage, au premier rang desquelles figurent *OpenAI*, *Microsoft*, *Meta* et *Google*, sont désormais confrontées à des interrogations quant à la durabilité technologique, financière, mais aussi environnementale de cette course à la consommation de ressources pour l'IA. Des investisseurs comme Eric Schmidt considèrent comme inéluctable l'augmentation de la consommation énergétique en raison des IA. Il prédit que les besoins en énergie et en matières premières seront tels qu'ils rendront inaccessibles les objectifs de lutte contre le réchauffement climatique. Il se repose sur l'espoir qu'à l'avenir, seules les IA seront à même d'aider à limiter les émissions de CO₂¹³².

Face à ces risques, de nouvelles architectures de *LLM* pourraient se développer dans les temps à venir. Ces architectures seraient à la fois décentralisées et de taille plus modeste. Comme le précise le *Wall Street Journal*, la course à la puissance ne semble pas adaptée à l'ensemble des besoins des entreprises.

« Un LLM [grand modèle de langage] géant entraîné sur l'ensemble des données du World Wide Web peut s'avérer vraiment excessif », déclare Robert Blumofe, directeur de la cybersécurité chez Akamai, société spécialisée dans la diffusion de contenus et le cloud computing. Pour les cas d'utilisation en entreprise, a-t-il ajouté, « vous n'avez pas besoin d'un modèle d'IA qui connaisse toute la distribution du 'Parrain', tous les films et toutes les émissions de télévision jamais

¹³¹ Gen Ai: Too Much Spend, Too Little Benefit? (Goldman Sachs Global Macro Research June 25, 2024) www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf

¹³² Former Google CEO Eric Schmidt Says 'We Are Never Going To Meet Our Climate Goals' – Pushes For AI And Data Centers To Solve The Crisis (Yahoo Finance, 10 Oct 2024) <https://finance.yahoo.com/news/former-google-ceo-eric-schmidt-161538746.html>

*réalisées. Oliver Parker, vice-président du marché mondial de l'IA générative chez Google Cloud, a déclaré qu'au cours des trois derniers mois, il avait vu des entreprises passer à des modèles de taille moyenne en partie parce que ces modèles répondent à des critères permettant de répondre à beaucoup plus de cas d'utilisation en entreprise. »*¹³³

Ainsi, des sociétés comme *OpenAI* commencent à décliner des versions plus modestes de leurs grands modèles et envisagent désormais de développer des solutions d'IA qui fonctionneront localement¹³⁴. Cela aura aussi pour avantage d'assurer aux sociétés qui emploieront ces IA, une meilleure maîtrise de leurs données. La protection des données qui servent à l'entraînement de ces modèles, mais aussi les données introduites par les utilisateurs eux-mêmes constituent en effet l'un des points faibles des services d'IA hébergés par les grandes sociétés. De nouveaux procédés cryptographiques sont désormais envisagés pour traiter les données de manière chiffrées dans les systèmes d'IA (suivant le principe du chiffrement dit *homomorphique*). L'objectif de ce champ de recherche est de protéger à la fois les données personnelles des utilisateurs et les données industrielles des entreprises. Il est à noter que l'Europe compte parmi les meilleurs talents mondiaux en cryptographie et pourrait développer à l'avenir de nouvelles générations de technologies qui associeront IA et sécurité cryptographique¹³⁵. Un autre champ de recherche qui devra être développé pour protéger les données

¹³³ “A giant LLM [large language model] that’s been trained on the entire World Wide Web can be massive overkill,” said Robert Blumofe, chief technology officer at cybersecurity, content delivery and cloud computing company Akamai. For enterprise use cases, he said, “You don’t need an AI model that knows the entire cast of ‘The Godfather,’ knows every movie that’s ever been made, knows every TV show that’s ever been made.” Oliver Parker, vice president of global generative AI go-to-market at Google Cloud, said he has seen enterprises shifting to midsize models in the last three months, in part because the models meet criteria capturing a lot more enterprise use cases.

These AI Models Are Pretty Mid. That’s Why Companies Love Them. Companies are looking for simpler and cheaper ways to deploy artificial intelligence (Isabelle Bousquette - Wall Street Journal, 17 July 2024) www.wsj.com/articles/these-ai-models-are-pretty-mid-thats-why-companies-love-them-710a0f72?mod=tech_lead_story

¹³⁴ OpenAI Slashes the Cost of Using Its AI With a “Mini” Model (Wired 18 Jul 2024) www.wired.com/story/openai-gpt-4o-mini/

¹³⁵ Paris-based Zama raises €67 million Series A to build privacy-preserving blockchain and AI (EU-Startups.com 7 Mar 2024) www.eu-startups.com/2024/03/paris-based-zama-raises-e67-million-series-a-to-build-privacy-preserving-blockchain-and-ai/

personnelles utilisées pour l'entraînement des IA concerne *"l'apprentissage fédéré"*, que la CNIL définit ainsi : « *L'apprentissage fédéré est un paradigme d'apprentissage dans lequel plusieurs entités entraînent collaborativement un modèle d'IA sans mise en commun de leurs données respectives. En pratique, les entités impliquées dans l'apprentissage envoient les modèles appris sur leurs données locales à un centre orchestrateur afin de consolider le modèle global. Ce paradigme s'oppose à l'apprentissage centralisé dans lequel toutes les données sont transmises à un serveur central chargé d'exécuter l'apprentissage du modèle*¹³⁶. »

Ainsi, dans le domaine des données sensibles, une meilleure protection des données de santé pourrait également passer par le développement des technologies d'IA sur des réseaux fédérés entre les acteurs locaux du secteur de la santé en Europe. Ces initiatives devraient permettre la valorisation scientifique et économique des données de santé, par les structures académiques ou les entreprises du secteur français et européen de la santé.¹³⁷

Un autre principe fondateur de l'Internet pourrait ainsi être appliqué au développement de l'IA à l'ère des réseaux ; il s'agit du principe de neutralité. Suivant le principe de neutralité de l'Internet, le réseau se comporte de manière neutre vis-à-vis des données qu'il transporte et l'intelligence du réseau se situe à ses extrémités (ce principe est aussi appelé architecture de « *bout en bout* » ou « *end to end* »).

Le principe de neutralité de l'Internet, promu par l'UE lors des travaux des Nations unies sur la régulation de l'Internet¹³⁸, pourrait aussi être appliqué à l'IA pour la puissance de calcul. Le magazine *The Economist* résume cela ainsi : « *L'Internet est devenu plus rapide et plus performant en rapprochant les données des utilisateurs.*

¹³⁶ Apprentissage fédéré (CNIL 2024) www.cnil.fr/fr/definition/apprentissage-federe

¹³⁷ C'est le cas d'*ADLIN Science*, start-up issue de l'École Polytechnique basée au Genopole d'Evry, et du projet *Alliance Santé IA* porté par le CHU de Montpellier, qui visent à développer de nouveaux usages de l'IA dans les structures de recherche et de soin.

¹³⁸ Communication from the Commission to the Council, the European Parliament, the European Economic and Social committee and the Committee of the Regions - Towards a Global Partnership in the Information Society : the Contribution of the European Union to the Second Phase of the World Summit on the Information Society (WSIS) <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52005DC0234>

Il doit désormais en être de même avec la puissance de calcul »¹³⁹. Des architectures dites distribuées pourraient également s'imposer pour la conception des IA. Parallèlement aux grands modèles de langage prônés par les acteurs du cloud comme un prolongement de leurs activités de traitement et de stockage des données, des IA plus « frugales » en ressources de calcul et en données pourraient alors se développer.

¹³⁹ "The internet got better and faster by moving data closer to users. Now the same must happen with computing power" (The Economist Technology Quarterly Jan 2024) www.economist.com/technology-quarterly/2024/01/29/the-internet-got-better-and-faster-by-moving-data-closer-to-users

X. IA ET PROPRIÉTÉ INTELLECTUELLE : NOUVELLES FRONTIÈRES JURIDIQUES

A. Licences et régulation du code des IA

Dans un premier temps, les bibliothèques d'IA utilisées pour l'apprentissage profond (*Deep Learning*) ont, dans leur grande majorité, été publiées en *open source* et pour nombre d'entre elles, ont été conçues par des Européens (c'est le cas de *Scikit-Learn*, *Keras* ou encore *SpaCy*). Avec l'essor des *grands modèles de langage (LLM)*, des questions nouvelles se posent aux acteurs de l'IA sur l'intérêt économique des formats ouverts pour les *LLM*. Ainsi, certaines sociétés ont fait le choix de conserver des licences *open source* pour leurs modèles, tandis que d'autres ont adopté des restrictions propriétaires sur leurs modèles les plus élaborés. Mais, au-delà du code lui-même, les données nécessaires à l'entraînement des modèles posent des problèmes spécifiques. Ainsi, Lawrence Lessig, spécialiste mondialement reconnu du droit de la propriété intellectuelle et avocat historique des logiciels libres, adopte une position nuancée lorsqu'il est question de l'ouverture des modèles d'IA :

L'IA est plus une catégorie qu'une technologie. Comme la catégorie « arme », elle va du relativement inoffensif au potentiellement catastrophique. Personne ne croit que l'accès que nous permettons aux sarbacanes devrait être le même que pour les missiles Stinger. De même, nous ne devrions pas croire que les normes logicielles développées pour les systèmes d'exploitation ou les lecteurs multimédias doivent s'appliquer de la même manière aux systèmes d'IA hautement performants qui ont le potentiel de causer des dommages considérables. Les entreprises privées seules, en compétition féroce les unes avec les autres, n'ont pas suffisamment d'incitations pour éviter ce risque catastrophique. Interdire simplement l'IA open source n'éviterait pas non plus le risque de dommages importants. Au lieu de cela, nous devons développer nos capacités réglementaires pour assurer un environnement dans lequel des IA sûres peuvent être développées, et notre jugement en matière de régulation pour déterminer quand le risque public de déploiement d'IA est trop grand.

*Aujourd'hui, ces risques sont imposés à tous par des acteurs privés peu soumis à une supervision publique. Cette formule n'a pas fonctionné avec les technologies dangereuses dans le passé. Elle ne fonctionnera pas avec les systèmes d'IA du futur.*¹⁴⁰

Cette logique de comparaison entre les IA et les technologies militaires conduit déjà certains experts à envisager des restrictions dans l'architecture même des puces graphiques (*GPU*) chargées du fonctionnement des modèles d'IA. En dehors des interdictions à l'exportation mises en place par le gouvernement américain, de nouvelles formes de restrictions technologiques pourraient aussi être envisagées.

Ainsi Will Knight, dans son article pour le magazine *Wired*,¹⁴¹ explore le concept de limitation des usages à risque pour l'IA en intégrant des contrôles directement dans les puces graphiques (*GPU*). Il s'agirait d'inscrire des règles dans l'architecture même de ces puces afin de limiter la puissance des algorithmes d'IA, et ainsi limiter le développement secret de systèmes d'IA potentiellement dangereux par des nations hostiles ou des entreprises peu scrupuleuses. Cette proposition avancée par le think tank américain *Center for a New American Security (CNAS)*, correspondrait à l'émission de licences par un régulateur pour contrôler l'accès aux plus hautes puissances de calcul nécessaire pour entraîner les IA les plus

¹⁴⁰ AI is more a category than a technology. Like the category “weapon”, it ranges from the relatively harmless to the potentially catastrophic. No one would believe that the access we allow to pea-shooters should be the same for stinger missiles. Neither should we believe that the software norms developed for operating systems or media players must apply in the same way to highly capable ai systems with the potential to cause immense harm. Private companies alone, in fierce competition with each other, do not have sufficient incentives to avoid catastrophic risk. Neither would simply banning open-source ai avoid the risk of great harm. Instead, we need to develop the regulatory capacity to ensure an environment within which safe ai can be developed, and the regulatory judgment to determine when the public risk from any ai deployment is too great. Today, these risks are imposed upon all of us by private actors with little public oversight. That formula has not worked with dangerous technologies in the past. It will not work with the ai systems of the future.

Not all AI models should be freely available, argues a legal scholar (Lawrence Lessig in *The Economist* 29 Jul 2024) www.economist.com/by-invitation/2024/07/29/not-all-ai-models-should-be-freely-available-argues-a-legal-scholar

¹⁴¹ Will Knight. Etching AI Controls Into Silicon Could Keep Doomsday at Bay (*Wired* 25 Jan 2024) www.wired.com/story/fast-forward-ai-silicon-doomsday/

puissantes. Ce concept rappelle la tentative avortée de l'administration Clinton d'installer une porte dérobée dans les puces des ordinateurs personnels (*Clipper chip*¹⁴²). Il repose sur l'idée que des composants sécurisés, comme les modules cryptographiques déjà présents dans certaines puces *Nvidia* et *Intel* pourraient restreindre l'utilisation non autorisée de certains modèles d'IA.

B. IA : quel impact sur les auteurs et les créateurs ?

Pour les auteurs et les créateurs, les IA génératives et la facilité donnée à des non-spécialistes d'élaborer des contenus (textes, images ou encore vidéos) par l'écriture de scripts (ou prompts) constituent un risque majeur et nécessiteront au-delà des textes existants de nouvelles mesures d'encadrement. Ainsi, comme le déclarent Jean-Marie Cavada et Colette Bouckaert dans les *Cahiers de la Documentation Française* :

*L'apparition de l'IA générative en 2022, la branche de l'IA qui se consacre à la création de modèles capables de générer de nouvelles données a été une déflagration pour l'ensemble des secteurs culturels. Ils ont vu leurs contenus pillés massivement, illégalement pour alimenter les bases d'entraînement de cette IA. On peut parler d'un choc, nécessitant la transformation de l'économie des industries culturelles.*¹⁴³

En effet, la rapidité avec laquelle les IA génératives se sont développées a pris de court les législateurs. Ainsi, la version initiale de l'*AI Act* ne prenait pas en compte les aspects liés au droit d'auteur. Cette prise en compte s'est avérée d'autant plus nécessaire que les acteurs des technologies perçoivent les revendications des auteurs comme des obstacles majeurs au développement des modèles

¹⁴² The Short Life and Humiliating Death of the Clipper Chip (Gizmodo 7 Apr 2023) gizmodo.com/life-and-death-of-clipper-chip-encryption-backdoors-att-1850177832

¹⁴³ Jean-Marie Cavada et Colette Bouckaert: Intelligence artificielle et propriété intellectuelle: quels progrès ? (Cahiers français, septembre-octobre 2024 N°441)

économiques des nouvelles générations de services. Une illustration de ce souci très relatif pour la légalité et le respect de la propriété intellectuelle a récemment été donnée par Eric Schmidt, l'ancien PDG de *Google*. Lors d'une récente conférence à *Stanford*, il conseillait aux étudiants de réaliser leurs projets d'IA en ces termes :

"Faites-moi une copie de TikTok, volez tous les utilisateurs, volez toute la musique, mettez-y mes préférences, produisez ce programme dans les 30 secondes qui suivent, publiez-le et dans une heure, s'il n'est pas viral, faites quelque chose de différent dans le même esprit. Donc, dans l'exemple que j'ai donné du concurrent de TikTok, et soit dit en passant, je ne disais pas qu'il fallait voler illégalement la musique de tout le monde. Ce que vous feriez si vous étiez un entrepreneur de la Silicon Valley, ce que vous serez tous, j'espère, c'est que si le produit décolle, vous engagerez tout un tas d'avocats pour aller nettoyer le désordre, n'est-ce pas ? Mais si personne n'utilise votre produit, peu importera que vous ayez volé tout le contenu. Et ne me citez pas..." [sic]¹⁴⁴

La devise « *bouger vite et casser des choses* » (*move fast and break things*) avait été formulée par Mark Zuckerberg pour expliquer aux développeurs mais aussi aux investisseurs les méthodes qui expliquent le succès de *Facebook*¹⁴⁵. Depuis lors, cette même devise a été adoptée par de nombreuses sociétés comme *Uber* ou *Amazon*. Ces entreprises ont associé à leurs nombreux avocats, des lobbyistes pour infléchir les régulations qui pouvaient nuire aux intérêts de ces sociétés. Dans le cas de *Facebook*, comme le révélait *The Guardian* en 2019, ces campagnes de lobbying pouvaient même prendre la forme d'intimidations ou de menaces auprès

¹⁴⁴ Make me a copy of TikTok, steal all the users, steal all the music, put my preferences in it, produce this program in the next 30 seconds, release it and in one hour, if it's not viral, do something different along the same lines. So in the example that I gave of the TikTok competitor, and by the way, I was not arguing that you should illegally steal everybody's music. What you would do if you're a Silicon Valley entrepreneur, which hopefully all of you will be, is if it took off, then you'd hire a whole bunch of lawyers to go clean the mess up, right? But if nobody uses your product, it doesn't matter that you stole all the content. And do not quote me.

Ex-Google CEO Schmidt advised students to steal TikTok's IP and 'clean up the mess' later (Fortune, 16 Aug 2024) fortune.com/2024/08/15/ex-google-ceo-eric-schmidt-stanford-ai-advice-steal-ip-hire-lawyers/

¹⁴⁵ Mark Zuckerberg's Letter to Investors: 'The Hacker Way' (Wired 1 Feb 2012) www.wired.com/2012/02/zuck-letter/

d'élus européens¹⁴⁶. Depuis la réélection de Donald Trump, les textes européens relatifs à la régulation des plateformes sont ouvertement contestés par les entreprises technologiques, soutenues par l'appareil d'État américain, qui les considère comme des entraves intolérables à leur développement économique, mais aussi comme des atteintes à l'expression politique de la puissance des États-Unis.¹⁴⁷

C. IA génératives : les nouvelles frontières du droit

Sur le plan juridique, les IA génératives posent de nombreuses questions nouvelles concernant l'usage de nouvelles formes de créations issues des technologies. En dehors même des sources utilisées pour entraîner ces IA et la nécessité de transparence et de négociations avec les ayants droit, des cas existeront où les résultats issus d'IA génératives pourront constituer une atteinte à la propriété intellectuelle. La procédure engagée par Scarlett Johansson contre *OpenAI* a ainsi rappelé qu'en dehors de cas impliquant des personnalités célèbres, il sera de plus en plus délicat de trancher des litiges dans lesquels une voix, une apparence ou encore une gestuelle auront été créées par une IA générative¹⁴⁸.

Un autre volet des questions liées à la propriété intellectuelle issues du développement des IA générative porte sur le domaine des inventions. Parallèlement au domaine de la création artistique, de nombreux inventeurs pourraient avoir recours à des IA génératives pour les aider à concevoir ou formaliser leurs travaux. Le *Wall Street Journal* s'inquiétait même des conséquences économiques que pourraient avoir les évolutions de la brevetabilité dans les années à venir s'il devenait impossible de protéger les inventions qui

¹⁴⁶ Revealed: Facebook's global lobbying against data privacy laws (The Guardian, 2 Mar 2019) www.theguardian.com/technology/2019/mar/02/facebook-global-lobbying-campaign-against-data-privacy-laws-investment

¹⁴⁷ EU rejects Trump's claim of unfair digital rules targeting US tech (Reuters, 27 Aug 2025) www.reuters.com/business/eu-rejects-trumps-claim-unfair-digital-rules-targeting-us-tech-2025-08-26/

¹⁴⁸ OpenAI v. Scarlett Johansson? Georgetown Law Professor Answers Legal Questions on AI-Generated Content (Georgetown University, 4 Jun 2024) www.georgetown.edu/news/ask-a-professor-openai-v-scarlett-johansson/

auraient été pour partie réalisées avec des IA : « *L'Office [américain] des brevets menace de rejeter toute demande de brevet dans laquelle il est considéré que l'inventeur a utilisé l'IA s'il est déterminé que l'invention n'a pas bénéficié d'une « contribution significative » d'une « personne physique ». Là encore, le terme « significatif » est utilisé de manière subjective et n'est jamais défini. Même s'il l'était, on peut se demander si les examinateurs pourraient évaluer avec précision l'importance de la contribution humaine*¹⁴⁹. »

Il y a deux ans, le journal *Le Monde* s'interrogeait sur la possibilité qu'un prix Nobel soit un jour attribué à une IA¹⁵⁰. Désormais, les IA sont étroitement associées aux travaux de recherche dans l'ensemble des disciplines scientifiques. À défaut de le décerner à une IA générale, le prix Nobel de physique a été attribué en 2024 à John Hopfield et Geoffrey Hinton, pionniers du développement des réseaux de neurones artificiels. Dans le même temps, le prix Nobel de chimie était attribué à Demis Hassabis et John Jumper, créateurs au sein de la société *Google DeepMind* de l'outil d'IA *AlphaFold* qui permet de prédire la conformation des protéines¹⁵¹.

¹⁴⁹ The patent office threatens to deny any patent application in which it is deemed that the inventor used AI if it is determined that the invention didn't have a "significant contribution" by a "natural person." There, too, the term "significant" is used subjectively and never defined. Even if defined, it is questionable whether examiners could accurately gauge the significance of the human contribution.

New Patent Guidance on AI Could Quash Innovation (Wall Street Journal, 11 Jul 2024)
www.wsj.com/articles/new-patent-guidance-on-ai-could-quash-innovation-dd848ea4

¹⁵⁰ L'intelligence artificielle, nouveau moteur de la recherche scientifique (Le Monde, 24 octobre 2022)
www.lemonde.fr/sciences/article/2022/10/24/l-intelligence-artificielle-nouveau-moteur-de-la-recherche-scientifique_6147154_1650684.html

¹⁵¹ AI comes to the Nobels: double win sparks debate about scientific fields (Nature, 10 Oct 2024)
www.nature.com/articles/d41586-024-03310-8

XI. LES DÉFIS ÉTHIQUES LIÉS AUX IA

A. Quelles valeurs humaines et politiques seront transmises aux AI ?

En l'espace de quelques années, la question de la confiance en l'IA est devenue centrale pour nos sociétés. À mesure que ces technologies sont associées à des

“ *Même les humains qui entraînent les réseaux profonds ne peuvent généralement pas regarder sous le capot et fournir des explications sur les décisions prises par leurs réseaux. La revue MIT Tech Review qualifie cette impénétrabilité de «sombre secret au cœur de l'IA»...*

Melanie Mitchell

fonctions essentielles dans l'ensemble des activités humaines, le besoin de les rendre à la fois intelligibles et prévisibles se fait plus pressant. L'une des premières questions posées sur le fonctionnement des IA est liée à leur imprévisibilité, y compris pour leurs concepteurs, comme le rappelle l'universitaire spécialiste de l'IA Melanie Mitchell : « *Même les humains qui entraînent les réseaux profonds ne peuvent généralement pas regarder sous le capot et fournir des explications sur les décisions prises par leurs réseaux. La revue MIT*

*Technology Review a qualifié cette impénétrabilité de « sombre secret au cœur de l'IA ». La crainte est que si nous ne comprenons pas comment fonctionnent les systèmes d'IA, nous ne pouvons pas vraiment leur faire confiance ou prédire les circonstances dans lesquelles ils commettront des erreurs*¹⁵². »

¹⁵² « Even the humans who train deep networks generally cannot look under the hood and provide explanations for the decisions their networks make. MIT's Technology Review magazine called this impenetrability 'the dark

Cette imprévisibilité des résultats issus des modèles d'IA rend leurs usages particulièrement délicats à intégrer dans des situations critiques pour la sécurité ou la liberté des personnes. Ainsi, pour Brian Christian, la nécessité d'intégrer les principes et valeurs d'une société dans les IA est d'autant plus grande que certains systèmes d'apprentissage automatique « *peuvent non seulement exprimer des préjugés, mais ils peuvent aussi les perpétuer de manière subtile et silencieuse*¹⁵³ ».

B. Les difficultés de « l'alignement » des IA

Les dilemmes moraux ou sociaux auxquels sont confrontés les utilisateurs, et surtout les concepteurs de ces IA, sont tels que les plus grands experts de l'IA, comme Stuart Russell, considèrent que le problème auquel nous sommes confrontés est lié à l'expression de nos objectifs lorsque nous interrogeons des IA.

*Trouver une solution au problème du contrôle de l'IA est une tâche importante. C'est peut-être, selon les mots du philosophe Nick Bostrom, « la tâche essentielle de notre époque ». Jusqu'à présent, la recherche en IA s'est concentrée sur des systèmes plus aptes à prendre des décisions, cela ne garantit pas pour autant que ces décisions seront meilleures lorsque les objectifs de l'homme et de la machine divergent. Ce problème nécessite un changement dans la définition de l'IA elle-même : d'un domaine qui s'intéresse à une notion unique d'intelligence comme optimisation d'un objectif donné à un domaine qui s'intéresse à la notion binaire de machines dont on pourra prouver ou non qu'elles sont bénéfiques pour les humains. Prendre le problème au sérieux permet de créer de nouvelles façons de penser l'IA, sa finalité et notre relation avec elle.*¹⁵⁴

secret at the heart of AI'. The fear is that if we don't understand how AI systems work, we can't really trust them or predict the circumstances under which they will make errors. »

Melanie Mitchell: Artificial Intelligence: A Guide for Thinking Humans (Pelican Books 2019)

¹⁵³ « Machine-learning systems like this not only demonstrate bias but may silently, subtly perpetuate it. »

Brian Christian. The Alignment Problem (W. W. Norton & Company 2020)

¹⁵⁴ Finding a solution to the AI control problem is an important task; it may be, in the words of philosopher Nick Bostrom, "the essential task of our age." Up to now, AI research has focused on systems that are better at making decisions, but this is not the same as making better decisions if human and machine objectives diverge.

En effet, les utilisateurs qui présument qu'une IA intégrera des considérations morales ou même des pratiques communément admises par les humains se heurtent à des réponses qui restent dénuées de ces préoccupations. Ce problème mène des IA à répondre à des questions d'une manière imprévue en suivant une « logique » qui reste fondamentalement étrangère à celle des êtres humains. Stuart Russell évoque ainsi la distance considérable qui peut exister entre les objectifs formulés par des humains et les résultats que leur apportent les IA :

C'est ce que nous appelons le problème du roi Midas. Le roi Midas a défini son objectif : « je veux que tout ce que je touche se transforme en or ». Il a obtenu exactement ce qu'il avait demandé. Malheureusement, cela comprenait sa nourriture, sa boisson et les membres de sa famille, et il est mort dans la misère et la faim. De nombreuses cultures racontent la même histoire. Le génie vous accorde trois vœux. Le troisième est toujours : « s'il vous plaît, annulez les deux premiers vœux », parce que j'ai détruit le monde. Et malheureusement, avec des systèmes plus intelligents et donc plus puissants que les nôtres, vous n'obtenez pas nécessairement un deuxième et un troisième vœu. Le problème vient donc de l'augmentation des capacités, couplée à notre incapacité à définir des objectifs complets et corrects. Pouvons-nous ramener notre dioxyde de carbone à des niveaux anciens afin de rétablir l'équilibre climatique ? Cela semble être un objectif formidable. Eh bien, le moyen le plus simple d'y parvenir est de se débarrasser de tous ces éléments qui produisent du dioxyde de carbone, qui se trouvent être des humains. Vous voulez guérir le cancer le plus rapidement possible. Cela semble formidable, n'est-ce pas ? Mais le moyen le plus rapide d'y parvenir est de mener des essais médicaux en parallèle avec des millions ou des milliards de sujets humains. Vous donnez donc le cancer à tout le monde et vous voyez ensuite quels traitements fonctionnent¹⁵⁵.

This problem requires a change in the definition of AI itself: from a field concerned with a unary notion of intelligence as the optimization of a given objective to a field concerned with a binary notion of machines that are provably beneficial for humans. Taking the problem seriously seems likely to yield new ways of thinking about AI, its purpose, and our relationship with it.

Stuart Russell: If We Succeed in Daedalus “AI & Society” Volume 151, Number 2; Spring by the American Academy of Arts & Sciences 2022)

www.amacad.org/sites/default/files/daedalus/downloads/Daedalus_Sp22_AI-and-Society.pdf

¹⁵⁵ We call this the King Midas problem. King Midas specified his objective: I want everything I touch turned to gold. He got exactly what he asked for. Unfortunately, that included his food and his drink and his family members, and he dies in misery and starvation. Many cultures have the same story. The genie grants you three

La tentation de présumer qu'un système d'IA, parce qu'il est basé sur des sources de données issues des humains, intégrera leurs préoccupations morales ou éthiques pourrait avoir des conséquences dangereuses pour les utilisateurs et, plus largement, pour les sociétés. L'objectif de permettre aux IA d'ordonner leurs priorités en fonction de préoccupations morales ou éthiques reste en effet l'un des principaux dilemmes non résolus par les concepteurs d'IA.

C. La confiance, objectif stratégique pour les concepteurs des IA

L'un des aspects stratégiques pour le développement des activités des spécialistes de l'IA sera lié à leur acceptabilité sociale et culturelle. À ce titre, la confiance des usagers sera déterminante pour la trajectoire industrielle de ces technologies. En 2014, dans leur ouvrage « *The Age of Context* », Scoble et Israel évoquaient déjà l'importance de la confiance pour le développement des services en ligne : « *Nous pensons que les entreprises les plus dignes de confiance seront à même de prospérer à l'ère du contexte, et celles qui manqueront de sincérité finiront par manquer de clients.* »

wishes. Always the third wish is “please undo the first two wishes” because I ruined the world. And unfortunately, with systems that are more intelligent and therefore more powerful than we are, you don’t necessarily get a second and third wish. So the problem comes from increasing capabilities, coupled with our inability to specify objectives completely and correctly. Can we restore our carbon dioxide to historical levels so that we get the climate back in balance? Sounds like a great objective. Well, the easiest way to do that is to get rid of all those things that are producing carbon dioxide, which happen to be humans. You want to cure cancer as quickly as possible. Sounds great, right? But the quickest way to do it is to run medical trials in parallel with millions of human subjects or billions of human subjects. So you give everyone cancer and then you see what treatments work.

AI could be a disaster for humanity. A top computer scientist thinks he has the solution. Stuart Russell interview by Kelsey Piper 26 Oct 2019 www.vox.com/future-perfect/2019/10/26/20932289/ai-stuart-russell-human-compatible

Transparence et fiabilité seront les facteurs de différenciation qui permettront aux clients de faire un nombre croissant de leurs choix. »¹⁵⁶

Les sociétés qui développeront les IA, devront s'appuyer sur des règles strictes pour garantir, dans la durée, que ces technologies ne représenteront pas des risques sociaux, culturels ou sécuritaires. C'est en montrant leur attachement à suivre ces règles que ces sociétés pourront devenir les architectes de la confiance vis-à-vis de leurs utilisateurs.

Ainsi, après ses recommandations sur la transparence qualifiée des algorithmes qu'il décrivait dans *The Black Box Society*¹⁵⁷, Frank Pasquale énonce, dans son ouvrage sur les nouvelles lois de la robotique, ce que devraient être les principes qui, à l'instar des règles proposées par Isaac Asimov, devraient organiser le fonctionnement des IA dans nos sociétés.

¹⁵⁶ We believe the most trustworthy companies will thrive in the Age of Context, and those found to be short on candor will end up short on customers. Transparency and trustworthiness will be the differentiating factors by which customers will make an increasing number of choices.

R. Scoble & S. Israel; Age of Context: Mobile, Sensors, Data and the Future of Privacy (CreateSpace Independent Publishing Sept 2014)

¹⁵⁷ Traduction française "Black Box Society: Les algorithmes secrets qui contrôlent l'économie et l'information" par Frank Pasquale (FYP Editions 2015) ouvrage original: "Black Box Society - The Secret Algorithms That Control Money and Information" by Frank Pasquale (Harvard University Press 2015)

Extrait de *New Laws of Robotics* par Frank Pasquale¹⁵⁸

- Les systèmes robotiques et l'IA doivent compléter les professionnels, et non les remplacer.
- Les systèmes robotiques et l'IA ne doivent pas contrefaire l'humanité.
- Les systèmes robotiques et l'IA ne doivent pas intensifier la course aux armements.
- Les systèmes robotiques et l'IA doivent toujours indiquer l'identité de leur(s) créateur(s), responsable(s) et propriétaire(s)

Les régulateurs devront donc exiger une responsabilité dès la conception pour compléter les modèles existants de sécurité dès la conception (*security by design*) et de protection de la vie privée dès la conception (*privacy by design*). Cela pourra impliquer d'exiger l'inscription dans le code, des audits ou des pratiques d'octroi de licences qui prendront en compte explicitement les résultats problématiques. De telles initiatives ne se contenteront pas de réglementer la robotique et l'IA a posteriori, mais influenceront le développement des systèmes en excluant certaines options de conception et en encourageant d'autres.¹⁵⁹

¹⁵⁸ Frank Pasquale. *New Laws of Robotics* (The Belknap Press of Harvard University Press 2020)

1. « Robotic systems and AI should complement professionals, not replace them. »
2. « Robotic systems and AI should not counterfeit humanity. »
3. « Robotic systems and AI should not intensify zero-sum arms races. »
4. « Robotic systems and AI must always indicate the identity of their creator(s), controller(s), and owner(s) »

¹⁵⁹ Regulators will need to require responsibility-by-design (to complement extant models of security-by-design and privacy-by-design). That may involve requiring certain hard-coded audit logs, or licensing practices that explicitly contemplate problematic outcomes. Such initiatives will not simply regulate robotics and AI post hoc, but will influence systems development by foreclosing some design options and encouraging others. Frank Pasquale. *New Laws of Robotics* (Harvard University Press 2020)

D. Des IA étroitement intégrées à la vie sociale, culturelle et politique

À partir du recueil en masse de données considérées comme « anodines » sur chaque individu, il devient possible de déduire des caractéristiques sur une personne qu'elle ne connaît pas elle-même ou qu'elle ne souhaite pas que l'on puisse deviner. Dès lors, la séparation entre des données dites « sensibles » et le reste des données personnelles devient pratiquement impossible à établir. Il a été établi que des habitudes de consommation pouvaient être associées à des choix politiques : En 1999, au moment de la crise de l'impeachment à l'encontre du président Clinton, le publicitaire Mark DiMassimo a analysé que « 84 % des consommateurs réguliers de soupes Campbell étaient favorables à la destitution du président Clinton¹⁶⁰ ».

Les IA seront aussi de plus en plus souvent associées à des décisions cruciales pour la vie, la sécurité et même la liberté des personnes. Ainsi, l'un des domaines les plus sensibles dans l'utilisation des données sur les individus concerne les statistiques criminelles et leur usage à des fins de prédiction des possibles récidives. Là encore, des boucles de rétroaction « toxiques » ont déjà été décrites dans l'usage de ces IA dans les systèmes judiciaires qui ont adopté ces technologies :

L'écart entre ce que nous souhaitons que notre outil mesure et ce que les données captent réellement devrait inquiéter les conservateurs comme les progressistes. Les criminels qui réussissent à échapper à une arrestation sont traités par le système comme des « individus à faible risque », ce qui incite à recommander la libération d'autres criminels similaires. Et les personnes surcontrôlées et condamnées à tort font partie du profil présumé des individus « à haut risque », ce qui incite le système à recommander la détention d'autres personnes comme elles. Ceci est particulièrement inquiétant dans le contexte de la police prédictive, où ces données de formation sont utilisées pour déterminer

¹⁶⁰ Darren Bridger et David Lewis : *Soul of the New Consumer: Authenticity What We Buy and Why in the New Economy* (Nicholas Brealey; Updated edition 2001)

*l'activité policière qui, à son tour, génère des données d'arrestation, créant ainsi une boucle de rétroaction potentielle sur le long terme.*¹⁶¹

Désormais, la prise en compte des conséquences potentiellement négatives de l'IA sur nos institutions et sur nos sociétés tout entières n'est plus uniquement le fait des techno-pessimistes inquiets des IA qui surpasseraient les humains et tenteraient de les asservir. Ainsi, Daron Acemoglu, prix Nobel d'économie 2024¹⁶² et professeur au *Massachusetts Institute of Technology* évoque même la nécessité d'établir un « principe de précaution » face au développement des IA dans nos sociétés :

*Même si je ne crois pas que la superintelligence et l'IA maléfique constituent des menaces majeures, je réfléchis souvent à la façon dont les risques actuels pourraient être perçus dans 50 ans. Le risque que nos enfants ou petits-enfants en 2074 nous accusent d'avancer trop lentement en 2024 au détriment de la croissance semble bien moindre que le risque que nous finissions par aller trop vite et détruire les institutions, la démocratie et bien plus encore dans le processus. Ainsi, les coûts des erreurs que nous risquons de commettre sont beaucoup plus asymétriques à la baisse. C'est pourquoi il est important de résister au battage médiatique et d'adopter une approche quelque peu prudente, qui peut inclure de meilleurs outils réglementaires, à mesure que les technologies de l'IA continuent d'évoluer*¹⁶³.

¹⁶¹ This gap, between what we intend for our tool to measure and what the data actually captures, should worry conservatives and progressives alike. Criminals who successfully evade arrest get treated by the system as “low-risk”—prompting recommendations for the release of other similar criminals. And the overpoliced, and wrongfully convicted, become part of the alleged ground-truth profile of “high-risk” individuals—prompting the system to recommend detention for others like them. This is particularly worrisome in the context of predictive policing, where this training data is used to determine the very police activity that, in turn, generates arrest data—setting up a potential long-term feedback loop.

Brian Christian. *The Alignment Problem, Machine Learning and Human Values* (p. 76) (W. W. Norton & Company 2020)

¹⁶² Three Win Nobel in Economics for Research on Global Inequality (New York Times, 14 Oct 2024) www.nytimes.com/2024/10/14/business/nobel-economics.html

¹⁶³ Although I don't believe superintelligence and evil AI pose major threats, I often think about how the current risks might be perceived looking back 50 years from now. The risk that our children or grandchildren in 2074 accuse us of moving too slowly in 2024 at the expense of growth seems far lower than the risk that we end up moving too quickly and destroy institutions, democracy, and beyond in the process. So, the costs of the mistakes that we risk making are much more asymmetric on the downside. That's why it's important to resist the hype

Un autre instrument de régulation des plateformes sera lié à l'analyse des algorithmes qui traitent les données personnelles. La transparence vis-à-vis du « Code » de ces algorithmes pourrait bientôt devenir un impératif pour les sociétés démocratiques. Cette régulation pourrait s'appliquer en particulier à la conception des algorithmes des objets connectés qui auront une influence sur la sécurité des personnes. C'est par exemple le cas pour les IA qui assurent le fonctionnement des véhicules autonomes. Ainsi, le *Media Lab* du MIT (*Massachusetts Institute of Technology*) a mis en place le projet *Moral Machine* pour analyser les choix éthiques et moraux des utilisateurs en cas d'accidents impliquant des voitures autonomes. Ce projet a permis de réaliser une enquête internationale auprès de personnes issues de 233 pays et territoires. L'analyse de cette enquête a permis de connaître les points de consensus mais aussi les divergences dans les choix éthiques des utilisateurs en fonction de leurs pays ou leurs cultures d'origine¹⁶⁴ : *"Les chercheurs ont constaté que les préférences des différents pays diffèrent considérablement, mais qu'elles sont également étroitement liées à la culture et à l'économie. Par exemple, les participants issus de cultures collectivistes comme la Chine et le Japon sont moins susceptibles d'épargner les jeunes que les personnes âgées – peut-être, selon les chercheurs, en raison d'une plus grande importance accordée au respect des personnes âgées ?"*¹⁶⁵

Les choix qui prévaudront au sein des IA qui auront des conséquences sur la vie et la sécurité des personnes, devront faire l'objet de débats démocratiques.

and take a somewhat cautious approach, which may include better regulatory tools, as AI technologies continue to evolve.

Interview with Daron Acemoglu professeur d'économie au Massachusetts Institute of Technology in Gen AI: Too Much Spend, Too Little Benefit? (Goldman Sachs Global Macro Research, 25 June 2024)

www.goldmansachs.com/intelligence/pages/gs-research/gen-ai-too-much-spend-too-little-benefit/report.pdf

¹⁶⁴ The Moral Machine Experiment (E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J-F.

Bonnefon & I. Rahwan - Nature, 24 Oct 2018) www.americaninno.com/wp-content/uploads/2017/05/The-MM-Experiment.pdf

¹⁶⁵ The researchers found that countries' preferences differ widely, but they also correlate highly with culture and economics. For example, participants from collectivist cultures like China and Japan are less likely to spare the young over the old—perhaps, the researchers hypothesized, because of a greater emphasis on respecting the elderly.

Should a self-driving car kill the baby or the grandma? Depends on where you're from (MIT Technology Review, 24 Oct 2018) www.technologyreview.com/2018/10/24/139313/a-global-ethics-study-aims-to-help-ai-solve-the-self-driving-trolley-problem/

En effet, ces choix s'ils nous étaient imposés par des acteurs extra-européens, pourraient avoir des conséquences sociales ou politiques qui s'inscriraient parfois à l'opposé des principes et des valeurs des citoyens européens. La transparence ne devra donc pas s'appliquer seulement au code de ces IA mais bien à leurs méthodes d'entraînement, afin que ces applications soient « alignées » à l'avenir sur les choix approuvés démocratiquement par les citoyens européens.

XII. LES ENJEUX GÉOPOLITIQUES DE L'IA

Au-delà des conséquences économiques, sociales ou même culturelles des IA, ces technologies représentent désormais un enjeu géopolitique. Elles interviennent directement dans la compétition économique et militaire entre les États-Unis, la Chine et l'Europe, mais aussi dans le cadre des tensions nouvelles entre la Russie et l'Union européenne. À l'ère des nouvelles guerres hybrides, les IA pourraient ainsi jouer un rôle crucial dans la redéfinition des rapports de force mondiaux en particulier avec les États-Unis. La facilité avec laquelle les technologies d'IA peuvent devenir des instruments d'aide à la décision ou à la mise en œuvre d'actions militaires fait de ces technologies un enjeu géopolitique, en particulier dans la confrontation entre les États-Unis et la Chine. Les technologies d'IA permettent désormais l'analyse en temps réel des zones de conflit, mais aussi le guidage des missiles, la coordination des essaims de drones ou encore la conception d'armes « intelligentes » entièrement autonomes¹⁶⁶. Ainsi, le double usage (civil et militaire) des technologies d'IA pourrait avoir d'importantes répercussions sur les rapports de force entre puissances dans les années à venir, ce qui faisait dire aux analystes de la *Rand Corporation* : « *Bien que les technologies aient souvent influencé la géopolitique, la perspective de l'IA signifie que les technologies elles-mêmes pourraient devenir des acteurs géopolitiques*¹⁶⁷. »

¹⁶⁶ Jonah M. Kessel, Natalie Reneau et Melissa Chan, « A.I. Is Making It Easier to Kill (You). Here's How », *The New York Times*, 13 Dec 2019 ; www.nytimes.com/video/technology/100000006082083/lethal-autonomous-weapons.html.

¹⁶⁷ Although technology has often influenced geopolitics, the prospect of AI means that the technology itself could become a geopolitical actor.

AI and Geopolitics : How Might AI Affect the Rise and Fall of Nations? (Barry Pavel, Ivana Ke, Michael Spirtas, James Ryseff, Lea Sabbag, Gregory Smith, Keller Scholl, Dominique Lumpkin (Rand Corporation, 3 Nov 2023) www.rand.org/pubs/perspectives/PEA3034-1.html

A. Des guerres pour l'IA aux guerres via l'IA

Parallèlement à leur contribution à de nouvelles formes de suprématie économique, les technologies d'intelligence artificielle ont déjà un impact crucial sur la conception des armements et la conduite des opérations militaires

“ *Bien que les technologies aient souvent influencé la géopolitique, la perspective de l'IA signifie que les technologies elles-mêmes pourraient devenir des acteurs géopolitiques...*

Barry Pavel

(simulations plus immersives, optimisation du traitement des données radars, sonars et coordination de la cyberdéfense...).

Avec la guerre en Ukraine, ces technologies ont démontré qu'elles étaient désormais devenues essentielles pour la conduite des opérations sur les théâtres de conflits¹⁶⁸. Un autre volet géopolitique de l'IA concerne les technologies qui permettent de renforcer le contrôle idéologique et politique des populations. De la

détection précoce de la dissidence sur les réseaux sociaux jusqu'au suivi et au contrôle des activités des personnes en temps réel via la reconnaissance faciale, l'IA constitue un élément clé pour assurer la pérennité des régimes autoritaires. Dans le même temps, les IA ont permis la « démocratisation » des opérations de désinformation et d'ingérence à grande échelle conçues pour affaiblir les démocraties. Il est ainsi devenu possible de mettre en œuvre de manière entièrement automatisée des campagnes de désinformation et de propagande qui touchent des millions d'individus.¹⁶⁹ Et ce, sans que les personnes visées ne

¹⁶⁸ Jeremy Wagstaff, « Une nouvelle forme de guerre », Fonds monétaire international, décembre 2023 www.imf.org/fr/Publications/fandd/issues/2023/12/Case-Studies-New-model-army-Jeremy-Wagstaff.

¹⁶⁹ M. J. Baniyas, « Inside Countercloud: A Fully Autonomous AI Disinformation System », The Debrief, 16 Aug 2023 ; <https://thedebrief.org/countercloud-ai-disinformation>.

puissent se rendre compte que les messages qu'elles reçoivent sont issus d'une IA. Ainsi, comme le rappelaient Henry Kissinger, Eric Schmidt et Daniel Huttenlocher dans leur ouvrage *L'Âge de l'IA*, nos démocraties ouvertes, dont les rouages économiques et politiques reposent massivement sur les technologies numériques, constituent des cibles de choix pour des entités étrangères hostiles ou des organisations terroristes¹⁷⁰.

B. États-Unis : ralentir mais pas étouffer les technologies d'IA chinoises

L'intelligence artificielle est ainsi devenue l'une des composantes majeures de l'affrontement entre les États-Unis et la Chine, ce que certains auteurs nomment désormais la « *guerre froide 2.0* ». Mais, au-delà du conflit économique, c'est la confrontation militaire directe entre les États-Unis et la Chine que certains appréhendent désormais. Cette confrontation pourrait avoir pour point de départ une attaque de l'île de Taïwan. C'est d'ailleurs à Taïwan que, dans leur immense majorité, les puces de haut niveau nécessaires au fonctionnement des IA sont produites par la société *TSMC (Taiwan Semiconductor Manufacturing Company)*.

Pour l'administration américaine, la crainte d'une suprématie chinoise sur l'IA est double : d'une part, que les technologies d'IA ne permettent à l'armée chinoise de rattraper son retard sur les technologies militaires américaines et, d'autre part, que ces IA ne permettent à la Chine d'accélérer son développement économique au point de dépasser les États-Unis. Ainsi, la première riposte américaine a porté, en octobre 2022, sur les investissements américains dans les sociétés chinoises d'IA ¹⁷¹. Ce que le *Center for Strategic and International Studies (CSIS)* analyse

¹⁷⁰ Henry Kissinger, Eric Schmidt et Daniel Huttenlocher, *The Age of AI: And Our Human Future* (Little, Brown and Company 2021)

¹⁷¹ « Biden bans range of US high tech investments in China citing national security risk », *The Guardian*, 9 Aug 2023 ; www.theguardian.com/world/2023/aug/09/biden-executive-order-us-investment-chinese-technology.

ainsi : *L'administration Biden dit essentiellement à la Chine : "Si votre politique est la fusion entre [technologies] militaires et civiles, le seul moyen réaliste de mettre en œuvre notre politique de non-usage militaire de ces ressources est de mettre fin à toutes ventes à la Chine, et nous sommes désormais prêts à franchir ce pas*¹⁷². »

Les États-Unis ont ainsi établi un quasi-embargo vis-à-vis de la Chine sur les technologies critiques pour la conception et le développement des IA : microprocesseurs de haut niveau, supercalculateurs et modèles d'IA. Les mesures de rétorsion décidées par la Chine ont alors pris la forme d'une limitation à l'exportation de certains métaux critiques indispensables à la fabrication de composants technologiques. L'embargo américain sur les puces de haut niveau constitue un défi stratégique pour le développement de l'IA en Chine. Cet embargo traduit également la crainte réelle d'une suprématie chinoise sur les technologies d'IA.

Les efforts des États-Unis pour empêcher leur concurrent de se procurer les puces les plus avancées sont perçus par les autorités chinoises comme des atteintes intolérables au développement économique du pays. Face aux risques de décrochage économique en matière d'IA, l'ambassadeur de Chine aux États-Unis, Xie Feng, a décrit la situation en termes sportifs : *« C'est comme si dans une compétition de natation on obligeait l'adversaire à porter un maillot de bain dépassé alors que l'on porterait soi-même un Speedo...*¹⁷³ ».

Avant même l'entrée en fonction du président élu Donald Trump, l'administration Biden avait pris le 2 décembre 2024 un nouveau volet de sanctions à l'égard de la Chine¹⁷⁴. Ces sanctions visaient à limiter l'accès des industriels chinois aux technologies nécessaires au développement de nouveaux modèles d'intelligence artificielle qui peuvent avoir également des applications militaires.¹⁷⁵ Ces mesures

¹⁷² Gregory C. Allen, « Choking off China's Access to the Future of AI », Center for Strategic and International Studies, 11 Oct 2022 www.csis.org/analysis/choking-chinas-access-future-ai.

¹⁷³ China warns of retaliation to US curbs on investment and chips » (Financial Times, 19 Jul 2023) www.ft.com/content/ad1350b9-0e4b-40a4-bb70-4c236e513e7a.

¹⁷⁴ Commerce Strengthens Export Controls to Restrict China's Capability to Produce Advanced Semiconductors for Military Applications (Bureau of Industry and Security, 2 Dec 2024) www.bis.gov/press-release/commerce-strengthens-export-controls-restrict-chinas-capability-produce-advanced

¹⁷⁵ Chip war ramps up with new US semiconductor restrictions on China (The Guardian 3 Dec 2024)

concernaient tant l'accès aux microprocesseurs qu'aux technologies logicielles et matérielles impliquées dans la conception de ces puces. En réponse, la Chine a annoncé de nouvelles restrictions sur les exportations de minéraux essentiels à la fabrication de puces et des équipements militaires, tels que le gallium, le germanium et l'antimoine.¹⁷⁶ Dès son arrivée au pouvoir, Donald Trump a pris de nouvelles mesures de sanctions contre les technologies chinoises « duales » utilisées à la fois à des fins civiles et militaires. Cependant, face à la montée en puissance des sanctions américaines sur la Chine, et en particulier la menace d'élévation des taxes douanières, le régime chinois a réagi en menaçant de limiter ses exportations de terres rares vers les États-Unis et aussi l'Union européenne. À l'issue de ces tensions, l'administration Trump a décidé en décembre 2025 d'autoriser l'exportation conditionnelle vers la Chine des puces *Nvidia* H200 dédiées à l'IA¹⁷⁷. Cette autorisation partielle est assortie d'un volet "transactionnel", en effet 25 % des revenus générés en Chine par *Nvidia* devront être reversés au gouvernement américain. Dans le même temps, malgré les demandes importantes de leurs entreprises technologiques, les autorités chinoises ont souhaité limiter les commandes effectuées auprès des fabricants de puces américains afin de permettre la montée en puissance de leurs propres technologies de puces et ainsi diminuer leur dépendance vis-à-vis des États-Unis ainsi que de Taïwan avec *TSMC*. L'Union européenne, quant à elle, envisage désormais la mise en place de mesures inédites de transferts technologiques dans le domaine de l'IA depuis la Chine vers l'Europe¹⁷⁸, pour autoriser les entreprises chinoises à accéder au marché européen¹⁷⁹.

www.theguardian.com/us-news/2024/dec/03/joe-biden-china-microchip-export-restrictions-law-changes

¹⁷⁶ La Chine interdit les exportations de certains minéraux essentiels vers les États-Unis (NHK 3 décembre 2024) www3.nhk.or.jp/nhkworld/fr/news/20241204_03/

¹⁷⁷ U.S.–China Chip Tensions Renew Focus on AI Controls as Washington Clears Conditional Nvidia Exports (FinTech Weekly, 12 Dec 2025) www.fintechweekly.com/magazine/articles/us-china-chip-controls-nvidia-h200-conditional-export-policy

¹⁷⁸ What Europe must learn from China's ascent in critical technologies (Bruegel, 2 July 2025) www.bruegel.org/newsletter/what-europe-must-learn-chinas-ascent-critical-technologies

¹⁷⁹ EU floats conditions such as tech transfers for China investments (Reuters, 15 Oct 2025) www.reuters.com/world/china/eu-floats-conditions-such-tech-transfers-china-investments-2025-10-14/

C. Des IA de plus en plus énergivores et pourtant cruciales pour le développement durable

Les IA sont particulièrement consommatrices en ressources de calcul et donc en énergie (durant les phases d'apprentissage de modèles d'IA mais aussi lors de l'exécution des requêtes), et sont aussi exigeantes en matières premières coûteuses pour la fabrication des puces graphiques (*GPU*) de haut niveau¹⁸⁰. Sécuriser l'énergie nécessaire au fonctionnement des IA est ainsi devenu un objectif stratégique pour les grands acteurs des technologies. L'importance du développement des infrastructures énergétiques nécessaires à alimenter les data centers consacrés à l'IA est telle, qu'au lieu de mesurer leur puissance en unités de calcul par seconde (en GigaFLOPs ou ExaFLOPs), comme c'est le cas pour les supercalculateurs, ces data centers sont mesurés en fonction de leur puissance électrique installée (en megawatt ou gigawatt). Ainsi, *Microsoft*, *Google* et *Amazon* investissent désormais dans des filières dédiées à l'usage de l'énergie nucléaire pour alimenter leurs IA¹⁸¹. Leur objectif est double : maîtriser les coûts de l'énergie et limiter le bilan carbone de leurs data centers. Cependant, le développement de nouvelles générations de "*petits réacteurs nucléaires modulaires*" (*SMR*) nécessite des investissements importants et ces réacteurs ne seront pas disponibles avant la prochaine décennie¹⁸².

Paradoxalement, ces technologies d'IA constituent également un élément important dans la trajectoire « bas carbone » des autres filières industrielles. L'IA est par exemple amenée à jouer un rôle dans la détection des gisements de matières premières critiques nécessaires à la production des énergies renouvelables. Comme le précise Jef Caers, professeur en sciences de la Terre à l'université *Stanford*, l'IA pourrait aussi occuper une place essentielle dans l'optimisation

¹⁸⁰ AI Power Consumption: Rapidly Becoming Mission-Critical (Forbes, 20 Jun 2024)

www.forbes.com/sites/bethkindig/2024/06/20/ai-power-consumption-rapidly-becoming-mission-critical/

¹⁸¹ Nuclear-Powered AI: Big Tech's Bold Solution or a Pipedream? (Wall Street Journal, 22 Oct 2024)

www.wsj.com/business/energy-oil/nuclear-power-artificial-intelligence-tech-bb673012

¹⁸² Google Backs New Nuclear Plants to Power AI (Wall Street Journal, 14 Oct 2024)

www.wsj.com/business/energy-oil/google-nuclear-power-artificial-intelligence-87966624

énergétique des processus de traitement de ces métaux ¹⁸³. Preuve supplémentaire du caractère polyvalent des modèles d'IA, ce sont des technologies d'IA développées dans le domaine militaire (en particulier dans le cadre de la lutte antiterroriste) qui ont par la suite donné naissance à des technologies à usage civil permettant de détecter les gisements de métaux rares.

La dépendance croissante de nos économies aux matières premières critiques est telle que le secrétaire général de *l'Organisation de coopération et de développement économiques (OCDE)*, Mathias Cormann, émettait un avertissement sur les conséquences écologiques des restrictions de ces exportations : « *Relever le défi de la neutralité des émissions de CO₂ nécessitera une nette montée en puissance de la production et des échanges internationaux de matières premières critiques. Les responsables de l'action publique doivent examiner de près la façon dont la concentration de la production et des échanges ainsi que le recours accru aux restrictions à l'exportation influent sur les marchés internationaux des matières premières critiques. Nous devons veiller à ce qu'aucune pénurie de matières ne nous empêche de tenir nos engagements climatiques* ¹⁸⁴. »

¹⁸³ « Can AI Help Us Go Green? », entretien avec Jef Caers, Techsequences, 19 Jul 2023 ; www.techsequences.org/podcasts/2023/07/can-ai-help-us-go-green.

¹⁸⁴ « Approvisionnements en matières premières critiques : les risques pour la transition verte », OCDE, 11 avril 2023 ; www.oecd.org/fr/presse/approvisionnement-en-matieres-premieres-critiques-les-risques-pour-la-transition-verte.htm.

D. Fin de la globalisation : vers le « friendshoring » ?

Les mesures d'embargo prises par les États-Unis ne font pas l'unanimité quant à leurs effets sur le long terme. Ainsi, pour Peter Wenninck ¹⁸⁵, ex-PDG de la société néerlandaise ASML, qui produit les machines nécessaires à la fabrication des microprocesseurs de haut niveau, « *Il y a 1,4 milliard de Chinois, dont beaucoup sont intelligents. Ils proposent des solutions auxquelles nous n'avions pas encore pensé. Vous les forcez [en raison de l'embargo] à devenir très innovants* ¹⁸⁶ ». Ce risque est aussi confirmé par Ben Buchanan, ancien conseiller spécial de la *Maison-Blanche* pour l'IA :

*Il est en effet possible que des contrôles stricts des exportations offrent un avantage à court terme, mais entraînent des inconvénients à long terme. Couper l'accès de la Chine aux puces, inciterait immédiatement la Chine à investir encore plus dans la construction d'une industrie nationale pour les microprocesseurs. En effet, les acheteurs chinois n'auraient alors plus accès aux puces étrangères, l'industrie que la Chine a tenté de construire (et jusqu'ici largement échoué) pourrait finalement prospérer. Les fabricants de puces américains et des pays alliés verraient également leurs revenus diminuer considérablement, ce qui pourrait les empêcher d'investir dans la recherche et le développement qui leur permettent de conserver leur avance technologique*¹⁸⁷.

¹⁸⁵ ASML CEO Peter Wennink is reported to have said that cutting off China will essentially drive innovation in China, meaning that China can become quite competitive at the cost of others. Speaking to media, he said, "There are 1.4 billion Chinese, many of them smart. They come up with solutions that we have not yet thought of. You force them to become very innovative.

AI Chips for China Face Additional US Restrictions (The Diplomat, 5 Apr 2004)

thediplomat.com/2024/04/ai-chips-for-china-face-additional-us-restrictions/

¹⁸⁶ Rajeswari Pillai Rajagopalan, « AI Chips for China Face Additional US Restrictions », The Diplomat, 5 Apr 2024 ; thediplomat.com/2024/04/ai-chips-for-china-face-additional-us-restrictions/.

¹⁸⁷ Controls on chips would be risky. There is a possibility that strong export controls might provide an advantage in the short term but lead to downsides in the long term. Cutting off China's access to chips would create an immediate incentive for China to invest even more in building a domestic industry. Given that Chinese buyers would in such circumstances have no access to foreign chips, the industry that China has been trying (and largely failing) to build just might finally flourish. American and allied chip companies would also see massive hits to their revenues, potentially constraining them from investing in the R&D that maintains their technological edge.

Il est à noter que l'Union européenne et les États-Unis ne partagent pas les mêmes objectifs stratégiques vis-à-vis de la Chine. Pour les États-Unis, il s'agit d'abord de limiter les ambitions des autorités chinoises dans les domaines militaire et économique. Des ambitions pour lesquelles les avancées de l'IA constituent un élément central. Pour les pays de l'Union européenne, il est davantage question d'un rééquilibrage des échanges avec la Chine. Ces échanges ayant souvent été considérés comme déloyaux dans le passé. Comme le rappelle Matthew Eitel du *Center for European Policy Analysis* : « *Bien que l'Union européenne étudie les moyens de renforcer son contrôle des exportations, l'accent qu'elle met sur le « retrait des risques » [de-risking] des chaînes d'approvisionnement en provenance de Chine est de plus en plus axé sur les pratiques commerciales « déloyales » de la Chine, et non sur la limitation des progrès technologiques de la Chine.* »¹⁸⁸

Une autre conséquence de la guerre en Ukraine correspond à la remise en cause des échanges économiques avec des pays qui constituent désormais de nouveaux risques géopolitiques pour les États-Unis. Cette nouvelle réalité géopolitique se traduit par de profondes modifications des règles qui prévalaient depuis l'effondrement du bloc soviétique. Celles-ci induisent désormais la remise en question du mouvement de globalisation économique amorcé depuis l'après-guerre avec la création de l'Accord général sur les tarifs douaniers et le commerce (*GATT*) en 1947, devenu en 1995 *l'Organisation mondiale du commerce (OMC)*. Cette reconfiguration du commerce mondial que l'ancienne secrétaire américaine au Trésor, Janet Yellen, nommait le « *friend-shoring* » (ou « *ally-shoring* »), par opposition à l'« *offshoring* » (ou délocalisation), et qu'elle résumait ainsi : « *Le “friend-shoring” signifie que [l'activité économique se construira] au sein d'un groupe de pays qui adhèrent fortement à un ensemble de normes et de valeurs liées à leur façon de [...] mener le système économique mondial, et nous devons renforcer nos liens avec*

Buchanan, Ben; Imbrie, Andrew. *The New Fire: War, Peace, and Democracy in the Age of AI* (p. 246). MIT Press 2022

¹⁸⁸ Although the European Union is exploring ways to strengthen its export controls regime, its emphasis on “de-risking” supply chains from China is increasingly focused on China’s “unfair” trade practices, not on limiting China’s technological advancement.

US China Tech Controls Face Problematic Diagnosis; Matthew Eitel (Center for European Policy Analysis 13 May 2024) <https://cepa.org/article/us-china-tech-controls-face-problematic-diagnosis/>

ces partenaires et travailler ensemble pour veiller à ce que nous puissions répondre à nos besoins en matériaux critiques ¹⁸⁹. »

Il est intéressant de noter que cette opposition entre grandes régions du monde qui devait initialement isoler les seuls adversaires des États-Unis pourrait aussi se faire aux dépens de la puissance américaine. Ainsi, à l'issue des taxes douanières qui lui ont été imposées par l'administration Trump, le principal partenaire commercial et allié historique des États-Unis, le Canada, vient de rejoindre l'initiative européenne *SAFE (Security Action for Europe)*. Ce programme de la Commission européenne lancé en 2025 est conçu pour renforcer l'industrie de défense européenne et il doit permettre des achats conjoints d'équipements militaires. Ce programme doit mobiliser 150 milliards d'euros via des prêts qui seront disponibles à l'ensemble des partenaires européens auquel s'ajoutera désormais le Canada, seul partenaire non européen de cette initiative¹⁹⁰.

Cette reconstitution de deux (voire plusieurs) blocs économiques antagonistes suscite des craintes auprès des économistes qui évoquent en premier lieu les risques d'une confrontation sino-américaine que les hauts responsables chinois, surnomment déjà le « rideau de silicone ». Ainsi, comme le rappelle Solveig Godeluck dans *Les Échos*, « *la division du monde commercial en deux blocs, les amis et les non-amis, aurait un coût économique non négligeable. Le commerce international serait pénalisé “dans la mesure où les pays en développement non alignés seront exclus de l'orbite du friend-shoring”* », souligne le *Center for Strategic and International Studies (CSIS)*. *L'Organisation mondiale du commerce (OMC) a d'ailleurs estimé que le commerce entre blocs réduirait d'environ 5 % le PIB mondial à long terme* ¹⁹¹. »

¹⁸⁹ « US Treasury Secretary Janet Yellen on the next steps for Russia sanctions and “friend-shoring” supply chains », Atlantic Council, 13 Apr 2022 ; www.atlanticcouncil.org/news/transcripts/transcript-us-treasury-secretary-janet-yellen-on-the-next-steps-for-russia-sanctions-and-friend-shoring-supply-chains.

¹⁹⁰ SAFE: member states endorse agreement on the participation of Canada (Council of the EU 19 Dec 2025) www.consilium.europa.eu/en/press/press-releases/2025/12/19/safe-member-states-endorse-agreement-on-the-participation-of-canada/

¹⁹¹ Solveig Godeluck, « Washington promeut un nouvel ordre commercial mondial avec le “friend-shoring” », *Les Échos*, 7 février 2023 ; www.lesechos.fr/monde/etats-unis/washington-promeut-un-nouvel-ordre-commercial-mondial-avec-le-friend-shoring-1904348.

E. Affrontement entre modèles politiques et technologiques américains, chinois... et européens

Selon Anu Bradford, professeure de droit international à l'université Columbia, « certains commentateurs décrivent l'Europe comme une *« victime de la guerre technologique entre les États-Unis et la Chine »* ou comme *« une colonie prise entre les États-Unis et la Chine »*¹⁹² », avec *« moins de pouvoir de négociation pour déterminer son propre destin numérique »* et donc obligée de *« faire un choix »* entre les États-Unis et la Chine¹⁹³ ». Mais, dans le même temps, une autre guerre a commencé : celle de l'influence « horizontale » des modèles de technologie promus par la Chine et les États-Unis. La confrontation entre ces deux pays pourrait à terme faire dévier le développement de l'IA dans de nombreux pays. Les pays démocratiques préfèrent alors une régulation issue des textes européens à l'adoption de technologies sous un intense contrôle étatique chinois ou de grandes plateformes américaines, avec leurs risques avérés de dérives. La plupart des pays démocratiques ont ainsi pu constater l'échec de l'autorégulation des grandes plateformes technologiques. Ces plateformes ont en effet été confrontées à de nombreux scandales qui ont altéré leur image auprès des démocraties. Ainsi, l'utilisation incontrôlée des données personnelles dans l'affaire *Cambridge Analytica*, l'usage de ces plateformes par des groupes extrémistes à des fins d'insurrection lors de l'assaut du *Capitole* à Washington, le 6 janvier 2021, ou les accusations de complicité de nettoyage ethnique en Birmanie portées contre *Facebook* ont profondément modifié la perception de ces plateformes par les responsables politiques.

L'influence européenne, qu'Anu Bradford nomme *« l'effet Bruxelles »*, traduit l'impact des régulations technologiques européennes au-delà des frontières de

¹⁹² Voir sur ce point le rapport de Mme Catherine Morin-Desailly, « L'Union européenne, colonie du monde numérique ? », remis au Sénat le 20 mars 2013 ; www.senat.fr/rap/r12-443/r12-443.html.

¹⁹³ Anu Bradford, *Digital Empires The Global Battle to Regulate Technology*, Oxford University Press, 2023.

l'Union. Il existe un effet de levier reconnu des normes européennes sur les lois élaborées dans le reste du monde sur le numérique. Dans son ouvrage, *Digital Empires*, Anu Bradford décrit l'obligation, y compris pour les États-Unis, de s'inspirer des formes européennes de régulation, en particulier pour l'IA, plutôt que de continuer à prendre les risques associés à l'absence de régulation des grandes plateformes. Les textes européens apparaissent plus protecteurs des libertés que les autres régulations élaborées dans le monde. Cela a été le cas notamment pour le règlement général sur la protection des données (avec le RGPD), dont les principes ont été repris dans plus de 17 pays et territoires dans le monde : de la Californie au Brésil, l'Inde et même la Chine. La loi chinoise sur les données personnelles (*PIPL*, pour *Personal Information Protection Law*), adoptée en août 2021, s'applique aux individus comme aux entreprises et même aux administrations de l'État chinois. Cependant des exceptions existent si les contraintes liées au traitement des données personnelles des citoyens chinois « venaient à empêcher les administrations d'exercer leurs missions statutaires¹⁹⁴ ».

À l'inverse, les responsables politiques des régimes autoritaires choisiront les technologies et les régulations chinoises, qui leur assureront un contrôle toujours plus strict sur leurs populations. En effet, en s'appuyant sur les technologies chinoises, comme celle du *Crédit Social*, les régimes autoritaires peuvent bénéficier de l'expérimentation la plus massive d'ingénierie sociale jamais menée dans le monde. Le *Crédit Social* a ainsi recours à l'utilisation des technologies d'IA pour la reconnaissance faciale, l'analyse des activités individuelles et le contrôle des individus via la notation de leur comportement économique, social et politique au sein de la société chinoise. Le *Crédit social* exploite ainsi l'ensemble des données recueillies sur l'activité des personnes et pourrait même à terme intégrer leurs profils génétiques¹⁹⁵.

¹⁹⁴ Jamie P. Horsley, « How will China's privacy law apply to the Chinese state? » (New America, 26 Jan 2021) www.newamerica.org/cybersecurity-initiative/digichina/blog/how-will-chinas-privacy-law-apply-to-the-chinese-state.

¹⁹⁵ Sui-Lee Wee, « China is Collecting DNA From Tens of Millions of Men and Boys, Using U.S. Equipment » (The New York Times, 17 Jun 2020) www.nytimes.com/2020/06/17/world/asia/China-DNA-surveillance.html.

F. L'Union européenne : citadelle démocratique assiégée par trois empires ?

Au-delà des actions de désinformation, d'ingérence et des cyberattaques menées par la Russie et la Chine, les pays de l'Union européenne doivent désormais prendre en compte de tensions industrielles et politiques nouvelles avec les États-Unis. En effet, pour la première fois depuis la création de l'UE, les trois empires (Russie, Chine et États-Unis) partagent un intérêt commun : l'affaiblissement politique, économique et militaire des États membres de l'Union européenne. Certains acteurs industriels, comme Elon Musk, ont même exprimé le souhait d'une désintégration de l'Union : « *L'UE devrait être abolie et la souveraineté rendue aux pays, afin que les gouvernements puissent mieux représenter leurs citoyens.* »¹⁹⁶ Il convient de noter que le PDG de X (anciennement Twitter) a formulé ces propos à la suite d'une amende infligée à sa société au titre du *Digital Services Act* (DSA).

Ces propos reprenaient, en les amplifiant, les positions hostiles à l'UE de l'administration Trump. Les autorités américaines ont en effet pris conscience que « *l'effet Bruxelles* » pourrait avoir des conséquences significatives pour les acteurs technologiques américains, en particulier si d'autres régions du monde adoptaient des régulations similaires pour encadrer l'IA et les plateformes numériques. Ainsi, l'ensemble des administrations américaines, en particulier le *Département d'État*, a reçu pour directive du président des États-Unis de freiner la mise en œuvre des principales régulations européennes concernant les plateformes numériques : *RGPD*, *DSA*, *DMA* et *AIA*. Le *Bureau du représentant américain au commerce* (USTR) a même déclaré : « *Si l'Union européenne persiste dans cette voie, les États-Unis n'auront d'autre choix que d'utiliser tous les moyens à leur disposition pour contrer ces mesures abusives. Parmi les mesures envisagées figurent des taxes et des restrictions sur les*

¹⁹⁶ x.com/elonmusk/status/1997279325876367719

entreprises de services. La même approche pourrait être adoptée à l'encontre des autres pays qui s'inspireraient de la stratégie européenne. »¹⁹⁷

Ces réglementations européennes représentent en effet un frein économique, mais aussi un obstacle à la réalisation des objectifs politiques de l'administration Trump. En premier lieu, il s'agit de l'alignement des régimes européens sur les orientations idéologiques des États-Unis et de l'affaiblissement des capacités de négociation de l'UE. La dépendance accrue de l'Union européenne aux technologies américaines est ainsi instrumentalisée par l'administration Trump pour influencer l'avenir politique de l'Union, comme cela a été explicitement énoncé dans la stratégie américaine de sécurité publiée par la Maison Blanche en décembre 2025.¹⁹⁸

Au-delà des mesures politiques ou de lobbying, s'ajoutent désormais des mesures *ad hominem*, tels que les refus de visa pour les États-Unis, visant les individus et organisations considérés comme hostiles aux intérêts américains. Cela a été le cas pour Thierry Breton, ancien commissaire européen, ainsi que pour plusieurs responsables d'ONG européennes œuvrant contre la désinformation et les ingérences sur les réseaux sociaux. Ces nouvelles mesures d'intimidation, en plus de leur caractère inédit, pourraient être accompagnées d'actions supplémentaires visant à priver des individus ou des organisations d'accès à certaines technologies ou services américains. Certains responsables européens craignent désormais que puisse être mis en place un « bouton d'extinction » (ou *kill switch*) pour certaines organisations jugées hostiles, voire pour des pays entiers en Europe, si l'administration Trump devait en décider ainsi¹⁹⁹.

¹⁹⁷ 'Trump Administration Targets Europe's Digital Laws as a Threat to Basic Rights and U.S. Business (Wall Street Journal, 15 May 2025) www.wsj.com/politics/policy/trump-administration-targets-europes-digital-laws-as-a-threat-to-basic-rights-and-u-s-business-20db1016

¹⁹⁸ National Security Strategy of the United States of America November 2025 (White House) www.whitehouse.gov/wp-content/uploads/2025/12/2025-National-Security-Strategy.pdf

¹⁹⁹ Europe's Growing Fear: How Trump Might Use U.S. Tech Dominance Against It (New York Times, 20 Jun 2025) www.nytimes.com/2025/06/20/technology/us-tech-europe-microsoft-trump-icc.html

G. Des IA encore plus propices à la surveillance de masse

De nouvelles générations d'objets et dispositifs connectés comme les automobiles connectées, les capteurs des villes intelligentes ou encore les objets connectés de santé, diversifient encore plus les canaux « d'extraction » d'informations personnelles. Mais, plus que d'accroître les quantités d'informations recueillies sur les individus, ces évolutions technologiques ont modifié la nature même des activités de surveillance. Ces nouveaux instruments ont ainsi créé des cycles courts entre le recueil des informations et leur utilisation à des fins d'ingérence, de modification des opinions ou des comportements. Au-delà de la lutte contre les activités criminelles, de lutte antiterroriste ou même d'espionnage, il devient possible d'agir sur les individus afin de modifier à leur insu leurs comportements et même leurs convictions. Ces messages politiques hyperciblés ont été utilisés par la société *Cambridge Analytica*, en lien avec les services de renseignement russes²⁰⁰, pour influencer les électeurs américains lors de la campagne présidentielle de 2016 ou les citoyens britanniques lors de la campagne sur le *Brexit*.

L'intelligence artificielle permet désormais une extension des technologies de surveillance au-delà des systèmes traditionnels qui nécessitaient d'importants moyens humains et logistiques pour recueillir et analyser des données sur l'ensemble des citoyens. Pour les régimes autoritaires et les dictatures, ces technologies représente un enjeu crucial pour assurer la pérennité de leurs régimes voire étendre leur influence comme rarement auparavant. Michael Kanaan, le responsable de l'intelligence artificielle au ministère de la Défense américain évoque même l'état d'inquiétude de l'état-major à l'issue du discours de Vladimir Poutine qui déclarait en 2017 à propos de l'IA : « *L'intelligence artificielle est l'avenir, non seulement pour la Russie, mais pour toute l'humanité. Elle s'accompagne d'opportunités colossales, mais aussi de menaces difficiles à prévoir.* »

²⁰⁰ Whistleblower : Cambridge Analytica shared data with Russia (Euractiv, 17 May 2018)

www.euractiv.com/section/global-europe/news/whistleblower-cambridge-analytica-shared-data-with-russia/

*Celui qui deviendra le leader dans ce domaine deviendra le maître du monde ».*²⁰¹ Pour les militaires américains, cet avertissement géopolitique constituait en quelque sorte un défi lancé à l'ensemble des acteurs technologiques : « *Poutine venait de déclarer, devant le monde entier, tout ce qui était en jeu. Sans laisser aucune place au malentendu, il assimilait la supériorité dans l'intelligence artificielle à la suprématie mondiale, à un pouvoir comparable à la domination économique, voire nucléaire. Il l'a dit pour le grand public, mais il avait en tête une intention politique. Quiconque deviendra le leader dans ce domaine deviendra le maître du monde* »²⁰².

Au-delà des seules technologies militaires traditionnelles, Vladimir Poutine avait aussi à l'esprit les perspectives de contrôle politique et idéologique que les technologies d'IA permettent aujourd'hui de développer. Ben Buchanan, évoque aussi un « *moment Spoutnik* » pour les autorités chinoises, lorsqu'en 2015 *AlphaGo*, l'intelligence artificielle de la société *DeepMind*, a surpassé les meilleurs joueurs de go chinois. Il établit un parallèle entre l'accélération du programme spatial américain à l'issue de la mise sur orbite par l'Union soviétique du premier *Spoutnik* en 1957 et la réaction des autorités chinoises qui, à la suite de la victoire d'*AlphaGo*, ont fixé comme un objectif stratégique la suprématie chinoise dans l'IA²⁰³.

AlphaGo a fait pour la Chine ce que Spoutnik a fait pour les États-Unis. Il a fait de l'IA un nouveau terrain de compétition géopolitique, avec des implications évidentes pour la guerre et la paix. Il a accéléré les investissements du gouvernement chinois dans l'IA, qui se chiffrent en milliards de dollars chaque

²⁰¹ Putin says the nation that leads in AI 'will be the ruler of the world' (The Verge, 4 Sep 2017.)

www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world

²⁰² « Now, for all to hear, Putin had just declared everything at stake. Without any room for misunderstanding, he equated AI superiority to global supremacy, to a strength akin to economic or even nuclear domination. He said it for public consumption, but it was rife with political purpose. Whoever becomes the leader in this sphere will become the ruler of the world. »

Michael Kanaan T-Minus AI : Humanity's Countdown to Artificial Intelligence and the New Pursuit of Global Power p17 (BenBella Books 2020)

²⁰³ China's 'New Generation Artificial Intelligence Development Plan' (2017) (translation by DigiChina - Stanford University, 1 Aug 2017) digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/

*année, et a poussé davantage d'étudiants à s'orienter vers les filières mathématiques et scientifiques correspondantes.*²⁰⁴

Un autre exemple illustre l'importance stratégique pour la Chine du développement des technologies d'intelligence artificielle. En 2021, le piratage à grande échelle des serveurs *Microsoft Exchange* a ainsi été attribué à la Chine²⁰⁵. Cependant, contrairement aux piratages "traditionnels", celui-ci n'avait pas pour but l'espionnage, mais bien l'utilisation des données issues de plusieurs centaines de milliers de sociétés occidentales à des fins d'entraînement des modèles d'IA chinois. Il ne s'agissait donc pas seulement d'obtenir des éléments de propriété intellectuelle ou même des secrets d'État mais bien d'améliorer les systèmes d'IA chinois. Cela permettait à la Chine d'obtenir un renforcement réciproque : accroître la connaissance sur les individus et les entreprises occidentales et améliorer les systèmes d'IA chinois qui, en plus de leur intérêt économique, permettront d'analyser plus finement les données permettant de mieux contrôler les populations.

À l'avenir, les objectifs économiques, politiques et militaires pourraient être de plus en plus étroitement mêlés dans les stratégies des États vis-à-vis des technologies d'IA. C'est en particulier le cas pour la Chine, pour laquelle l'IA représente un instrument essentiel dans sa stratégie de leadership politique et économique mondial.

²⁰⁴ AlphaGo, the argument went, had done for China what Sputnik did for the United States. It established AI as a new terrain of geopolitical competition, one with obvious implications for war and peace. It accelerated Chinese government investment in AI to the tune of billions of dollars each year and drove more students into the relevant fields in math and science.

Buchanan, Ben; Imbrie, Andrew. *The New Fire: War, Peace, and Democracy in the Age of AI* (p. 51) (MIT Press 2022)

²⁰⁵ China's Microsoft Hack May Have Had A Bigger Purpose Than Just Spying (NPR Aug 26, 2021) www.npr.org/2021/08/26/1013501080/chinas-microsoft-hack-may-have-had-a-bigger-purpose-than-just-spying

Maîtriser l'IA, un objectif triplement stratégique pour la Chine

- pour le développement économique de l'ensemble de son secteur technologique et plus largement l'ensemble de son économie, dans le cadre de sa compétition pour le leadership mondial avec les États-Unis,
- pour maintenir un contrôle idéologique et politique sur sa population et donc être en mesure d'assurer la pérennité de son régime,
- pour développer les prochaines générations de systèmes d'armes intelligents dont les conflits récents (et en particulier la guerre menée par la Russie en Ukraine) ont montré l'importance stratégique.

H. L'IA au cœur du conflit entre la Chine et Taïwan

Un exemple de l'importance croissante des technologies d'IA dans la conduite des opérations militaires est donné par George Takach dans son ouvrage *Cold War 2.0* sur les nouvelles formes de conflictualité basées sur les technologies d'IA. Il y décrit ce que pourraient être les premiers instants de la guerre que la Chine mènerait contre Taïwan et l'importance stratégique que pourraient avoir les technologies d'IA dans ce conflit :

Il est fortement probable que le jour fatidique où la Chine lancera sa première salve de missiles et de drones sur Taïwan, l'IA se mettra rapidement en mode automatique. Pourquoi ? Pour la simple raison qu'il n'y aura pas assez de temps pour permettre aux militaires taïwanais et américains de s'insérer

efficacement dans les activités de sélection et d'approbation des cibles. Les yeux et les oreilles humains ne peuvent pas assez rapidement absorber de données dans ce scénario de bataille, et le cerveau humain ne peut pas digérer et traiter les informations assez vite pour savoir rationnellement quoi faire dans le cas où des centaines de missiles et de drones se dirigeront vers eux, la plupart atteignant leurs cibles prévues quelques minutes après le lancement. Ironiquement, l'ancien handicap du chef militaire, le brouillard de la guerre, où le chef militaire en savait trop peu sur ce qui se passait sur le champ de bataille, a été remplacé dans ce scénario d'attaque sur Taïwan par « l'écrasante clarté de la guerre » tout aussi problématique (le chef militaire a toutes les données, mais ne sait pas quoi en faire en temps voulu). La seule chance de succès pour les démocraties est que l'IA qui sous-tend tous les systèmes de défense aérienne effectue son travail de manière spectaculairement efficace et qu'elle le fasse de manière autonome²⁰⁶.

Les conflits récents en particulier en Ukraine démontrent que l'utilisation de l'IA sur le champ de bataille apporte des avantages cruciaux. Ces technologies pourraient ainsi être déterminantes lors de conflits à venir. À ce titre, la formation des spécialistes de l'IA et le développement des filières industrielles dans ce domaine deviennent des enjeux stratégiques pour l'ensemble des États. George Takach emprunte même à Winston Churchill sa métaphore sur la *Royal Air Force*²⁰⁷ à l'appui de sa démonstration. Mais cette fois ce n'est plus aux pilotes de

²⁰⁶ Chances are, on that fateful day when China launches the opening shower of missiles and drones on Taiwan, the AI will quickly be set to automatic mode. Why? For the simple reason that there will be insufficient time to permit Taiwanese and American military personnel to effectively insert themselves into target selection and approval activities. Human eyes and ears cannot meaningfully absorb sufficient data points in this battle scenario quickly enough, and the human brain cannot digest and process information fast enough to know rationally what to do in the case of hundreds of missiles and drones coming at them, most of which will hit their intended targets a few minutes after launch. Ironically, the previous handicap of the military leader, the so-called fog of war (where the leader knew too little of what was going on in the battle space), has been replaced in this Taiwan attack scenario by the equally problematic “overwhelming clarity of war” (the leader has all the data points, just doesn’t know what to do with them all in a timely manner). The only chance of success here for the democracies is that the AI underpinning all the different defensive ADS weapons does its job spectacularly well, autonomously.

George S. Takach. Cold War 2.0_ Artificial Intelligence in the New Battle between China, Russia, and America by George S. Takach (p. 15) (Pegasus Books 2024)

²⁰⁷ 'Never in the field of human conflict was so much owed by so many to so few' (UK Parliament, 20 Aug 1940) www.parliament.uk/about/living-heritage/transformingsociety/private-lives/yourcountry/collections/churchillexhibition/churchill-the-orator/human-conflict/

Spitfire, mais aux spécialistes des technologies d'IA que les nations démocratiques pourraient devoir leur liberté :

La victoire dans ce que l'Histoire appellera la bataille du détroit de Taïwan sera déterminée par la maîtrise dans chaque camp de 4 technologies essentielles : l'intelligence artificielle, les semi-conducteurs, les ordinateurs quantiques et la biotechnologie. Le succès (ou l'échec) de ces quatre technologies dépendra de 2 facteurs : d'une part, de la qualité des diplômés en sciences, technologie, ingénierie et mathématiques (STEM) qui ont suivi des programmes de maîtrise et de doctorat dans les autocraties et les démocraties, respectivement, au cours des vingt-cinq années précédant la bataille, et d'autre part, de la façon dont les complexes militaro-industriels des 2 camps auront intégré ces innovations dans leurs systèmes d'armes et leurs écosystèmes de défense respectifs. Fusionner l'innovation civile et militaire est un travail très difficile. Jamais auparavant autant n'aura dépendu d'un si petit nombre de diplômés universitaires²⁰⁸

En plus des revendications traditionnelles de la Chine envers sa « province rebelle », l'île de Taïwan est aussi le siège de la première entreprise mondiale de production de puces de haut niveau : *TSMC (Taiwan Semiconductor Manufacturing Company)*. Ainsi, malgré des tentatives récentes de reproduire dans des usines aux États-Unis, en ou en Europe les activités de *TSMC*, la production des usines basées à Taïwan restent cruciales, en particulier pour les puces nécessaires au fonctionnement des IA. Cet assaut contre Taïwan, s'il était couronné de succès, représenterait donc une triple victoire pour le régime chinois :

²⁰⁸ Victory in what history will call the Battle of the Taiwan Strait will be determined by the degree of each side's mastery over four critical technologies: artificial intelligence, semiconductor chips, quantum computers, and biotechnology. In turn, the success (or failure) of these four technologies will depend on two factors: first, the quality of the science, technology, engineering, and mathematics (STEM) graduates that have been coming out of master's and PhD programs in the autocracies and the democracies, respectively, over the twenty-five years prior to the battle, and second, how well the military-industrial complex in both camps has integrated these innovations into their respective weapons systems and defense ecosystems. Fusing civilian and military innovation is very hard work. Never before will so much depend on so few college grads.

George S. Takach. Cold War 2.0_ Artificial Intelligence in the New Battle between China, Russia, and America (p. 17). (Pegasus Books 2024)

- victoire politique : cette réunification forcée, longtemps annoncée par le régime, renforcerait auprès des citoyens chinois le « narratif » du régime sur le nationalisme et l'importance de la puissance militaire chinoise,
- victoire économique : la possibilité pour la Chine d'inverser à son avantage les restrictions technologiques imposées par les États-Unis, ce qui ralentirait l'ensemble des acteurs technologiques occidentaux et permettrait à la Chine de se rapprocher du leadership mondial dans le domaine des technologies d'intelligence artificielle,
- victoire géostratégique : par la validation de la stratégie des autorités chinoises visant à contrôler l'ensemble de la mer de Chine, ainsi que ses voies maritimes, par l'extension des implantations successives sur l'ensemble des îlots en mer de Chine.

XIII. POLITIQUE INDUSTRIELLE DE L'IA : LES LEVIERS DE LA PROTECTION DES DROITS DE L'HOMME

A. Vers une « troisième voie » éthique pour les IA européennes

Par sa logique d'encadrement des pratiques à risque en matière d'IA, *l'Artificial Intelligence Act (AI Act)* a pour objectif d'éviter l'évolution vers un contrôle autoritaire des populations comme celui mis en place par le *Crédit social* en Chine. *L'AI Act* favorisera également le développement des compétences en éthique de l'IA au sein des entreprises et administrations, et aussi à terme d'activités économiques liées aux IA éthiques. Pour Mark Coeckelbergh, philosophe spécialiste des aspects éthiques de l'IA, l'introduction des règles éthiques dès la conception des modèles d'IA pourrait même constituer un avantage concurrentiel pour les technologies européennes.

D'une manière générale, l'éthique de l'IA ne consiste pas nécessairement à interdire des choses (comme le souligne Boddington en 2017). Un autre obstacle à la mise en pratique de l'éthique de l'IA est que de nombreux acteurs du domaine de l'IA, tels que les entreprises et les chercheurs, considèrent encore l'éthique comme une contrainte, comme quelque chose de négatif. Cette idée comporte une part de vérité : souvent, l'éthique impose des limites et définit ce qui est inacceptable. Et si nous prenons l'éthique de l'IA au sérieux et mettons en œuvre ses recommandations, nous pourrions devoir faire certains compromis, en particulier, à court terme. L'éthique peut avoir un coût : en termes d'argent, de temps et d'énergie. Toutefois, en réduisant les risques, l'éthique et l'innovation responsable soutiennent le développement durable et à long terme des entreprises et de la société. Démontrer à l'ensemble des acteurs du domaine de l'IA que c'est le cas restera difficile, en particulier pour les décideurs politiques. Il convient également de noter que la politique et la

réglementation ne consistent pas seulement à interdire ou à rendre les choses plus difficiles ; elles peuvent également apporter un soutien, en mettant en place par exemple des mesures d'incitations²⁰⁹.

“ Il n’y a pas une seule technologie de l’iPhone qui n’ait pas été financée par l’État fédéral américain. Cela inclut les technologies des réseaux sans fil, l’Internet, le GPS, l’écran tactile et, plus récemment, l’assistant personnel à commande vocale Siri...”

Mariana Mazzucato

sans fil, l’Internet, le GPS, l’écran tactile et, plus récemment, l’assistant personnel à commande vocale Siri²¹⁰. »

À l’avenir, les politiques industrielles de l’Union européenne pourraient aussi être liées à la perception de l’urgence à devenir une puissance politique et militaire face

Un autre aspect crucial des politiques industrielles en matière de technologies sera lié à une meilleure articulation entre recherches civiles et militaires. En effet, de nombreuses technologies fondamentales de l’Internet ont d’abord été développées dans le domaine militaire. Ainsi, comme le rappelle l’économiste Mariana Mazzucato : « *Il n’y a pas une seule technologie de l’iPhone qui n’ait pas été financée par l’État fédéral américain [en particulier par la recherche militaire]. Cela inclut les technologies des réseaux*

²⁰⁹ That being said, generally speaking AI ethics is not necessarily about banning things (Boddington 2017). Another barrier to getting AI ethics to work in practice is that many actors in the AI field such as companies and technical researchers still think of ethics as a constraint, as something negative. This idea is not totally misguided: often ethics has to constrain, has to limit, has to say that something is unacceptable. And if we take AI ethics seriously and implement its recommendations, we might face some trade-offs, in particular in the short term. Ethics may have a cost: in terms of money, time, and energy. However, by reducing risks, ethics and responsible innovation support the long-term, sustainable development of businesses and of society. It is still a challenge to convince all the actors in the AI field, including policymakers, that this is indeed the case. Note also that policy and regulation are not only about banning things or making things more difficult; they can also be supportive, offering incentives, for example.

Mark Coeckelbergh. AI Ethics (pp. 174-176) (MIT Press 2020)

²¹⁰ Mariana Mazzucato, citée par Amanda Schaffer, « Tech’s Enduring Great-Man Myth », MIT Technology Review, 4 août 2015 ; www.technologyreview.com/2015/08/04/166593/techs-enduring-great-man-myth.

à des blocs potentiellement hostiles, comme ce fut le cas aux États-Unis après l'avertissement lancé par Vladimir Poutine à propos du rôle géopolitique de l'IA.

En effet, après avoir théorisé l'importance stratégique de l'intelligence artificielle, Vladimir Poutine est désormais conscient des risques que le retard de la Russie dans le domaine de l'IA fait peser sur son économie et surtout sur ses capacités militaires et politiques. En effet, en dépit des progrès réalisés par la Russie en matière de drones et d'armes autonomes depuis le déclenchement de la guerre en Ukraine, les autorités russes ont conscience de leur retard dans le domaine de l'IA vis-à-vis des États-Unis et de la Chine. Or, ce retard pourrait constituer dans les années à venir un handicap stratégique majeur pour la Russie. Cette situation pourrait expliquer la présence, dans le plan de paix en Ukraine proposé par les États-Unis et la Russie en novembre 2025, de dispositions relatives aux coopérations technologiques dans le domaine de l'intelligence artificielle en contrepartie du développement avec la Russie de projets d'extraction de terres rares dans l'Arctique. Ainsi, la disposition 13 de ce plan prévoyait : « *La Russie sera réintégrée à l'économie mondiale : la levée des sanctions sera discutée et convenue par étapes et au cas par cas. Les États-Unis concluront un accord de coopération économique à long terme pour le développement mutuel dans les domaines de l'énergie, des ressources naturelles, des infrastructures, de l'intelligence artificielle, des centres de données, des projets d'extraction de terres rares dans l'Arctique et d'autres opportunités commerciales mutuellement avantageuses. La Russie sera invitée à réintégrer le G8.* »²¹¹.

²¹¹ Extrait de "The Unfinished Plan for Peace in Ukraine: Provision by Provision"(Center for Strategic & International Studies November 24, 2025) www.csis.org/analysis/unfinished-plan-peace-ukraine-provision-provision

B. Quelle réponse européenne face à la techno (ou data) colonisation ?

Lorsque les Britanniques, les Russes et les Japonais ont tenté d'atteindre l'hégémonie aux XIXe et XXe siècles, ils s'appuyaient sur des bateaux à vapeur, des locomotives et des mitrailleuses. Au XXIe siècle, pour dominer une colonie, il n'est plus nécessaire d'envoyer des canonnières. Il faut extraire les données. Quelques entreprises ou gouvernements qui récolteraient les données du monde entier pourraient transformer le reste du globe en colonies de données – des territoires qu'ils contrôleraient non pas par la force militaire ouverte mais par l'information.

Yuval Noah Harari²¹²

Face au constat d'une Europe sur le point de devenir une « colonie numérique » de deux continents²¹³, des propositions de réponses en termes de politique industrielle commencent à être envisagées. La dépendance vis-à-vis des acteurs américains, en particulier dans le domaine du cloud où *Amazon (AWS)*, *Microsoft Azure* et *Google Cloud* totalisent à eux trois près des 75 % du marché du cloud en Europe, rend d'autant plus nécessaire le renouveau des stratégies des États européens dans le domaine de l'IA.

Ainsi, l'ancien président de la *Banque centrale européenne (BCE)*, Mario Draghi dans son rapport sur la compétitivité européenne, dresse un constat sévère sur la dépendance et le retard de l'UE dans le domaine des technologies : " *Nous comptons sur une poignée de fournisseurs pour les matières premières essentielles*

²¹² When the British, Russians and Japanese made their bids for hegemony in the nineteenth and twentieth centuries, they relied on steamships, locomotives and machine guns. In the twenty-first century, to dominate a colony, you no longer need to send in the gunboats. You need to take out the data. A few corporations or governments harvesting the world's data could transform the rest of the globe into data colonies – territories they control not with overt military force but with information.

Harari, Yuval Noah. Nexus (p. 370). Vintage Publishing 2024

²¹³ « Le "technocolonialisme" vise à contrôler les technologies qui sous-tendent l'économie mondiale »

(tribune d'Hermann Hauser et Hazem Danny Nakib parue dans Le Monde du 26 août 2024)

www.lemonde.fr/idees/article/2024/08/26/le-technocolonialisme-vise-a-controler-les-technologies-qui-sous-tendent-l-economie-mondiale_6295347_3232.html

et nous importons plus de 80 % de nos technologies numériques. Nous sommes sévèrement en retard dans les nouvelles technologies : seulement 4 des 50 plus grandes entreprises technologiques mondiales sont européennes...²¹⁴”.

Mario Draghi considère qu'il est urgent pour l'UE d'investir collectivement le champ des technologies stratégiques en particulier dans le domaine de l'IA²¹⁵. Pour cela, il préconise une action commune d'une ampleur inédite pour que l'Europe ne reste pas uniquement dans une posture défensive vis-à-vis des États-Unis et de la Chine. Jusqu'à une période récente, le principe d'un endettement commun, que l'on nommait parfois « *Eurobonds* », se heurtait au refus systématique de nombreux interlocuteurs européens et en particulier allemands. La crise liée à la pandémie de Covid-19 et plus récemment la guerre en Ukraine ont fait évoluer cette attitude. Ainsi, en 2020, l'UE a mis en place le plan de relance européen pour un montant initial de 750 milliards d'euros²¹⁶. Ce plan était destiné à la fois à lutter contre la pandémie (d'où le nom initial de « *Coronabonds* » donné à cet endettement) et dans un second temps à financer la transition énergétique et numérique des pays de l'Union.

Dans la continuité de ces mesures exceptionnelles d'endettement, le rapport élaboré par Mario Draghi propose d'instaurer, sur le long terme, un investissement européen destiné à financer le développement des technologies stratégiques pour l'UE, au premier rang desquelles figure l'IA. L'écart d'investissement entre les États-Unis et la Chine et les États de l'Union européenne n'a cessé de se creuser dans ces domaines. À titre d'illustration, depuis 2013, les États-Unis ont consacré à l'IA des investissements privés correspondants à 335,2 milliards de dollars, suivis par la Chine avec 103,7 milliards de dollars et du Royaume-Uni avec 22,3 milliards de dollars. Pour la seule année 2023, les investissements privés américains

²¹⁴ Address by Mr. Draghi – Presentation of the report on the Future of European competitiveness – European Parliament – Strasbourg – 17 September 2024

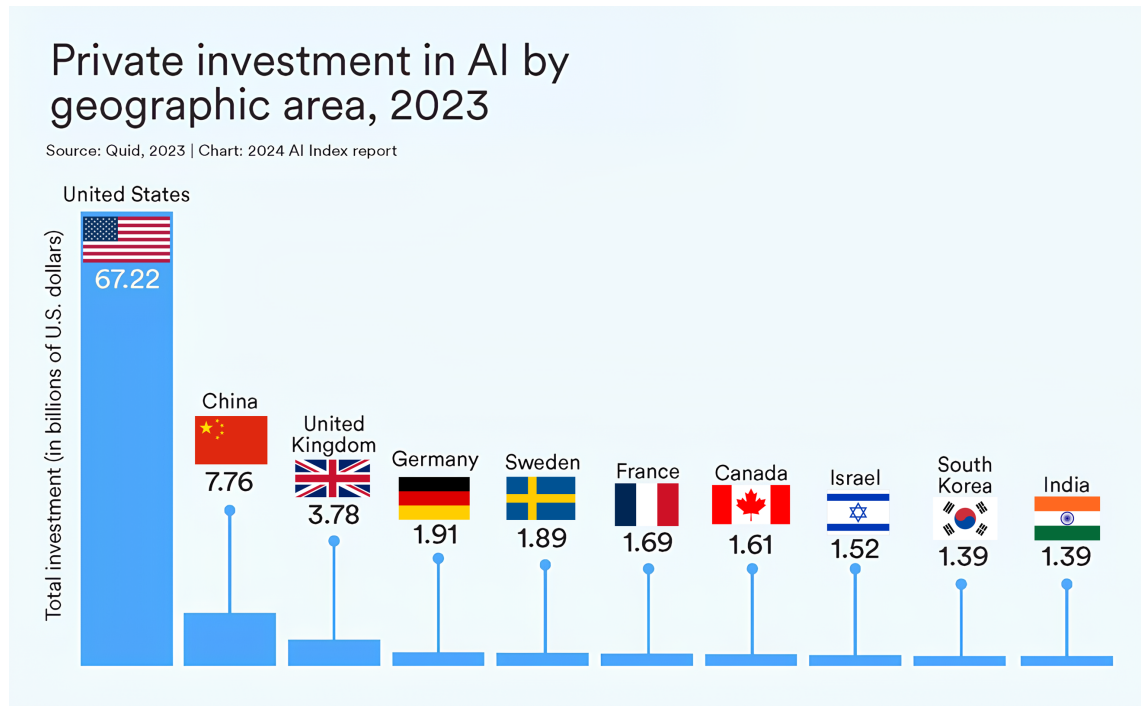
https://commission.europa.eu/document/download/fcbc7ada-213b-4679-83f7-69a4c2127a25_en

²¹⁵ The future of European competitiveness – A competitiveness strategy for Europe (Report by Mario Draghi Sept 2024) https://commission.europa.eu/topics/strengthening-european-competitiveness/eu-competitiveness-looking-ahead_en

²¹⁶ Plan de relance européen : où en est-on ? (Touteurope.eu 05 juin 2024)

www.touteurope.eu/fonctionnement-de-l-ue/plan-de-relance-europeen-queles-sont-les-prochaines-etapes/

représentaient 67,22 milliards de dollars quand les investissements allemands et français représentaient respectivement 1,91 et 1,69 milliards de dollars.



Private investment in AI by geographic area, 2023 - AI Index
Source : HAI Human-Centered Artificial Intelligence (Stanford University)²¹⁷

Pour Mario Draghi, afin d'être en mesure de rattraper son retard face aux États-Unis et à la Chine, l'Union européenne doit consentir à des investissements « massifs et sans précédent » : entre 750 et 800 milliards d'euros par an, soit près de 5 % du PIB européen.

Là encore, des méthodologies nouvelles devront être employées pour financer les technologies stratégiques pour l'Union européenne. De plus, seul un volontarisme politique fort sera à même de convaincre l'ensemble des pays de l'Union de mettre en place cet investissement commun²¹⁸. La situation géopolitique de l'Union

²¹⁷ <https://hai.stanford.edu/news/ai-index-state-ai-13-charts>

²¹⁸ Pour l'ancien président de la Banque centrale européenne (BCE), l'Union européenne (UE) doit réaliser à l'avenir 800 milliards d'euros d'investissements supplémentaires par an – l'équivalent de 5 % de son produit

européenne, désormais attaquée par trois empires, et menacée par la montée en puissance des populismes, rend ces efforts financiers et politiques plus nécessaires et moins inenvisageables que par le passé.

Mais, au-delà de la maîtrise par les acteurs européens des technologies cruciales pour leur développement économique, c'est aussi la protection du modèle social et démocratique européen qui est en jeu. À elles seules, les mesures juridiques d'encadrement des technologies d'IA ne pourront pas constituer une protection suffisante face aux risques de dérives de technologies qui seront présentes dans l'ensemble des aspects de la vie sociale, économique et politique.

Parallèlement à la régulation des plateformes d'IA existantes, les acteurs industriels européens devront être en mesure de développer des technologies qui respecteront les principes et valeurs des citoyens européens. Les acteurs industriels européens pourraient ainsi être à même de développer des alternatives aux services existants qui ne seraient pas basés sur l'extraction massive de données des utilisateurs à des fins publicitaires.

Pour participer pleinement aux évolutions des IA afin qu'elles soient à la fois durables et protectrices des libertés, l'Europe doit ainsi instaurer une politique industrielle d'une ampleur inédite. Au-delà des textes de régulation des technologies, seule une démarche commune d'investissement permettra de développer en Europe de futures générations de technologies éthiques fondées sur l'IA. Cela nécessitera de la part des Européens un effort encore plus important que celui qui a été fourni lors de la pandémie de Covid-19. Là encore, l'urgence liée aux crises géopolitiques déjà ouvertes en Europe et l'imminence possible d'autres conflits dans le monde rendront plus nécessaire encore ce volontarisme industriel européen tout particulièrement dans le domaine de l'IA.

intérieur brut (PIB) –, soit environ trois fois le plan Marshall (entre 1 % et 2 % du PIB en investissements annuels dans l'après-guerre).

Thomas Piketti : « Le rapport Draghi a l'immense mérite de tordre le cou au dogme de l'austérité budgétaire » (Le Monde 14 sept 2024) www.lemonde.fr/idees/article/2024/09/14/thomas-piketty-le-rapport-draghi-a-l-immense-merite-de-tordre-le-cou-au-dogme-de-l-austerite-budgetaire_6317119_3232.html

S'il n'existe pas d'alternatives européennes pour les technologies stratégiques, les mesures européennes d'encadrement des technologies de l'IA ne sauront, à elles seules, maîtriser les risques politiques, sociaux et issus de plateformes extra-européennes. Ainsi, la défense des droits humains devra aller de pair avec une politique industrielle qui permettra de concevoir et développer en Europe des IA qui intégreront dès l'origine les principes et valeurs des citoyens européens. Parallèlement à la mise en place de règles éthiques pour les technologies, il convient de développer une politique industrielle pour l'essor de l'IA. En effet, trop souvent, l'UE s'est préoccupée des risques liés à ces technologies sans se donner les moyens (politiques mais aussi économiques) de développer des alternatives aux technologies et aux services extra-européens.

Alors qu'il était commissaire européen chargé du marché intérieur, de la politique industrielle et du numérique, Thierry Breton avait déclaré qu'il convenait d'en finir avec la naïveté qui avait marqué l'action de l'Europe dans le domaine de la politique industrielle sur les technologies. Cette naïveté portait en premier lieu sur le rééquilibrage des échanges industriels et commerciaux avec les États-Unis et la Chine, dont les marchés sont encore fermés à la plupart des entreprises européennes. Ces déclarations effectuées en 2020 ont connu un surcroît d'actualité depuis la victoire de Donald Trump et ses mesures de rétorsion contre les industriels de l'Union européenne sous la forme de nouvelles barrières douanières. Une UE que Trump compare désormais à une "mini-Chine"²¹⁹ [sic].

Ainsi Thierry Breton, avant d'être interdit d'accès aux États-Unis par l'administration Trump, mentionnait un autre aspect de la naïveté européenne. Cette fois, son inquiétude portait sur les actions de déstabilisation politiques et de campagnes de désinformation menées en Europe par des entités ou de nations hostiles : « *Enfin, il nous incombe de protéger nos démocraties contre le fléau de la désinformation. Soyons clairs : des puissances étrangères bien identifiées mènent, sur notre continent, des opérations de déstabilisation de nos démocraties, de nos processus électoraux, de nos économies. Les grandes plateformes, au mieux par*

²¹⁹ Trump's trade tariffs: how protectionist US policies will hit German carmakers (The Guardian, 9 Nov 2024) www.theguardian.com/business/2024/nov/09/trumps-trade-tariffs-how-protectionist-us-policies-will-hit-german-carmakers

*manque d'anticipation, au pire par négligence, deviennent acteurs par procuration de ces stratégies.*²²⁰. »

Cependant, de nombreuses affaires ont démontré la difficulté voire l'impossibilité actuelle de l'Union européenne à contraindre de grandes plateformes comme *X* (ex-*Twitter*), *Facebook* ou plus encore *TikTok* à se conformer au droit européen. Qu'il s'agisse des ingérences d'Elon Musk dans les processus électoraux européens²²¹, de l'absence de transparence algorithmique de *TikTok*, ou des dérives liées à l'absence de protection des données personnelles dans le cas de *Facebook* avec l'affaire *Cambridge Analytica* ou plus récemment des modifications de la posture de Mark Zuckerberg vis-à-vis de la régulation des contenus de Facebook à l'issue de l'élection de Donald Trump. Dans le passé, les sanctions infligées par l'UE pour abus de position dominante de ces plateformes n'avaient pas modifié leurs comportements industriels et éthiques. Les mêmes questions d'efficacité des mesures de l'UE se posent lorsqu'il est question du comportement fiscal de ces sociétés, du respect du droit du travail ou plus généralement de la protection des données personnelles de leurs usagers.

En l'absence d'acteurs européens de taille internationale dans le domaine des technologies, notamment dans le domaine de l'IA, les mesures de régulation européennes resteront handicapées par des difficultés politiques et industrielles. C'est le cas lorsqu'il est question de remettre en cause l'hégémonie des trois principaux acteurs américains dans le domaine du cloud (*Amazon*, *Microsoft* et *Google*) en Europe. Ces plateformes sont en effet largement utilisées par les plus grandes entreprises européennes. Ces difficultés seront plus grandes encore lorsqu'il est question de mesures d'interdiction qui peuvent être politiquement délicates, comme en témoigne le cas de *TikTok* aux États-Unis. La loi votée par le *Congrès américain* en avril 2024 devait contraindre *ByteDance*, la société mère de *TikTok*, à vendre la branche américaine de *TikTok* à une société américaine, sous

²²⁰ La fin de la naïveté (tribune de Thierry Breton parue dans Les Echos le 10 août 2020)

www.lesechos.fr/idees-debats/cercle/la-fin-de-la-naivete-1229485

²²¹ EU Commission urged to act over Elon Musk's 'interference' in elections (The Guardian 8 Jan 2025)

www.theguardian.com/world/2025/jan/08/eu-commission-urged-to-act-elon-musk-interference-elections

peine d'interdiction du service aux États-Unis²²². Cette tentative d'interdiction a fait l'objet d'un arbitrage politique au plus haut niveau de l'État américain. Ainsi, Donald Trump qui était, en 2020, favorable à une interdiction de *TikTok* a confirmé qu'il s'opposerait à l'interdiction de cette plateforme dès son investiture le 20 janvier 2025. Il reprenait à cette occasion l'argumentaire des responsables de la plateforme phare des technologies chinoises au nom des « *risques faramineux que cette interdiction ferait peser sur la liberté d'expression aux États-Unis...* » [sic]²²³. En décembre 2025, la société *TikTok* a finalement annoncé que la branche américaine de *TikTok* passerait sous le contrôle d'actionnaires américains²²⁴.

C. Une synergie nécessaire entre politique industrielle et régulation des technologies

Dans le domaine des lois antitrust, l'UE a toujours choisi comme critère premier la lutte contre les monopoles intra-européens et donc les risques d'augmentation des prix pour le consommateur. Ce choix faisait passer au second plan les risques en matière de limitation des innovations liées aux pratiques monopolistiques. Cependant, l'UE pourrait être amenée à modifier sa doctrine afin de renforcer les acteurs européens des technologies face à leurs compétiteurs américains ou chinois. En effet, comme le précise Philippe Aghion, Prix Nobel d'économie 2025, pour ses recherches sur le rôle de la "destruction créatrice" dans le progrès

²²² Support for a U.S. TikTok ban continues to decline, and half of adults doubt it will happen (Pew Research Center - 5 Sep 2024) www.pewresearch.org/short-reads/2024/09/05/support-for-a-us-tiktok-ban-continues-to-decline-and-half-of-adults-doubt-it-will-happen/

²²³ TikTok says US ban would have 'staggering' impact on free speech (BBC 24 Sep 2024) www.bbc.com/news/articles/c5y3y79llndo

²²⁴ TikTok se plie aux exigences de Washington et crée une entreprise américaine pour éviter son interdiction (Le Monde 19 décembre 2025) www.lemonde.fr/pixels/article/2025/12/19/tiktok-se-plie-a-washington-et-cree-une-entreprise-americaine-pour-eviter-son-interdiction-sur-le-territoire_6658601_4408996.html

technologique : « *Au nom de la concurrence [au sein de l'UE], il n'y a pas eu de politique industrielle en Europe...* »²²⁵.

L'UE devra là encore effectuer un arbitrage de haut niveau entre les risques liés aux monopoles extra-européens et ceux qui seraient liés à l'émergence de géants européens. Il s'agira d'effectuer un choix entre des sociétés européennes plus susceptibles de répondre aux mesures de régulation, ou des géants américains ou chinois qui, en plus de leur hégémonie technologique, présentent des risques de dérive politique, d'atteintes aux libertés ou même d'ingérences. L'expérience européenne a montré que non seulement ces sociétés résistent aux mesures de régulation, mais qu'elles parviennent aussi à les contourner et à maintenir leur hégémonie malgré les sanctions infligées dans le cadre des lois antitrust de l'UE. Désormais ces grandes entreprises américaines bénéficient du soutien de l'ensemble de l'appareil d'État américain qui utilise tous ses leviers pour empêcher les législations européennes d'entraver leur développement au sein du marché européen des technologies.

Pour être en mesure de protéger les droits des Européens, les actions de régulation de la concurrence et les politiques industrielles européennes devront donc être décidées et coordonnées au plus haut niveau. Ces politiques devront en effet s'exercer en synergie pour aider à développer des acteurs européens de taille internationale qui seront seuls à même de répondre aux principes et valeurs des citoyens de l'UE.

Ben Buchanan, ancien conseiller spécial de la *Maison-Blanche* pour l'IA, rappelle le rôle crucial de la puissance publique américaine dans le développement industriel des IA. Qu'il s'agisse de faciliter l'accès des PME à la commande publique *via* une administration dédiée (la *Small Business Administration*), d'offrir des conditions d'accès privilégiées aux larges capacités de calcul nécessaires au fonctionnement des IA ou encore d'aider à constituer de larges sources de données accessibles aux entreprises pour l'entraînement des IA.

²²⁵ Affaires étrangères (France Culture 3 janvier 2026) www.radiofrance.fr/franceculture/podcasts/affaires-etrangeres/2026-l-europe-face-aux-populismes-7609287

Les gouvernements démocratiques devraient rendre la puissance de calcul plus accessible aux personnes qui ont de grandes idées. Les gouvernements pourraient acheter des crédits de cloud computing à grande échelle et à bas prix auprès d'entreprises leaders. Ils pourraient ensuite accorder l'accès à ces ressources informatiques à des laboratoires universitaires ou à des start-ups offrant de larges avantages sociaux, à l'image de la National Science Foundation et de la Small Business Administration qui offrent des subventions aux chercheurs et aux entrepreneurs. En plus de favoriser l'innovation, rendre les ressources informatiques plus accessibles est essentiel à l'ère de l'IA, et ce d'autant plus que de nombreux algorithmes qui préservent la confidentialité ou qui sont indépendants des données nécessitent d'importantes puissances de calcul pour fonctionner²²⁶.

Pour être en mesure de faire prévaloir les principes et valeurs des Européens, l'Union européenne a constitué un bloc de textes de régulation qui encadrent l'activité des grandes plateformes de l'Internet. Ces textes ont vocation à limiter les multiples dérives possibles liées à l'activité de ces plateformes (protection des données personnelles, lutte contre les abus de position dominante, modération des contenus préjudiciables, interopérabilité et transparence algorithmique). Ces textes représentent des initiatives pionnières et ont donné lieu à des déclinaisons dans de nombreux autres pays au-delà des frontières de l'UE, comme cela a été le cas pour le *RGPD (règlement général sur la protection des données)*. Les règlements récents *DSA (Digital Services Act)* et *DMA (Digital Markets Act)*, *DGA (Data Governance Act)* et plus récemment l'*AI Act*, constituent un ensemble de réponses aux risques de dérives de ces plateformes.

Cependant, l'absence de solutions alternatives européennes pour les citoyens et les entreprises constitue un frein politique à la mise en œuvre des mesures les plus

²²⁶ Democratic governments should make computing power more accessible to people with big ideas. Governments could buy cloud computing credits from leading companies at a large scale and at low rates. They could then grant access to those computing resources for academic labs or startups with broad social benefits, much as the National Science Foundation and Small Business Administration offer monetary grants to researchers and entrepreneurs. In addition to increasing innovation, making computational resources more accessible is essential for the age of AI, especially since many of the privacy-preserving or data-independent algorithms described above require substantial computational power to work.

Buchanan, Ben; Imbrie, Andrew. *The New Fire: War, Peace, and Democracy in the Age of AI* (p. 239). MIT Press 2022

fortes. Comme le démontre la difficulté à établir des mesures d'interdiction de *TikTok* aux États-Unis ou en Europe, malgré les risques de manipulations et d'ingérence dans les processus démocratiques.

La perception d'une Europe qui serait uniquement défensive et arc-boutée sur la régulation face aux acteurs technologiques américains avait déjà été résumée de manière lapidaire par Barack Obama en 2015 : « *Nos entreprises ont créé l'Internet, elles l'ont développé, perfectionné d'une manière telle que les Européens n'arrivent pas à rivaliser. Et ce que l'on dépeint souvent comme une vision élevée des Européens sur certaines questions [comme la protection de la vie privée] n'est en fait que le moyen qu'ils ont trouvé pour privilégier leurs propres intérêts commerciaux*²²⁷. »

D'une manière symétrique pour la Chine, Jack Ma, le patron du géant chinois *Alibaba*, déclarait à *VivaTech Paris* en 2019, sans cacher son ironie, ses doutes face aux règles mises en place par les Européens dans le domaine numérique : « *Je suis inquiet des inquiétudes de l'Europe. Vous avez des règles de droit si parfaites. Tout le monde doit suivre ces règles, et si certains s'inquiètent, alors ils créent de nouvelles règles de droit. Mais nous, quand on a un problème [en Chine], on commence par résoudre le problème*²²⁸. »

Quelques mois plus tard, Jack Ma avait l'audace de critiquer publiquement la politique financière de Xi Jinping. Il était alors l'homme le plus riche de Chine et sa compagnie s'apprêtait à effectuer l'introduction en Bourse la plus importante de tous les temps²²⁹. Et, ironiquement, c'est l'absence de règles liées à l'État de droit en Chine (et en particulier l'absence d'indépendance de la justice) qui a permis qu'il soit assigné à résidence, puis évincé *d'Alibaba*, la compagnie qu'il avait fondée et

²²⁷ "We have owned the internet. Our companies have created it, expanded it, perfected it in ways that they can't compete. And oftentimes what is portrayed as high-minded positions on issues sometimes is just designed to carve out some of their commercial interests."

Obama accuses EU of attacking American tech companies because it 'can't compete' (The Verge 15 Feb 2015) www.theverge.com/2015/2/17/8050691/obama-our-companies-created-the-internet

²²⁸ Intervention de Jack Ma à VivaTech, Paris le 23 mai 2019
<https://x.com/AlibabaGroup/status/1131673464773914624>

²²⁹ Jack Ma's Costliest Business Lesson: China Has Only One Leader (Wall Street Journal, 20 Aug 2021) www.wsj.com/articles/jack-mas-costliest-business-lesson-china-has-only-one-leader-11629473637

enfin l'a forcé à l'exil sans manquer d'être soumis à une rééducation idéologique pour faire l'apologie du régime chinois...

Il convient de noter qu'à chaque fois que le régime chinois a eu à choisir entre le développement économique des entreprises technologiques chinoises et le respect de la ligne idéologique du régime, il a toujours préféré l'orthodoxie idéologique. Cela au point, comme le décrit le *Financial Times*, de prendre le risque d'étrangler la totalité du secteur des technologies en Chine²³⁰. Comme le souligne le *Wall Street Journal*, le développement des IA chinoises doit désormais épouser les formes politiques et idéologiques du régime dans lesquelles elles sont conçues.

*La Chine a pris de l'avance dans la révolution de l'IA en développant des systèmes capables de voir et d'analyser le monde à très grande vitesse. Le domaine de l'IA connu sous le nom de vision par ordinateur, qui permet le suivi et la surveillance, s'inscrit dans la lignée de l'accent mis par le dirigeant chinois Xi Jinping sur le contrôle politique. Malgré ce succès initial, le pays a été pris de court par le lancement public de ChatGPT d'OpenAI fin 2022 et l'engouement pour l'IA générative qu'il a déclenché. Les grands modèles de langage de l'IA générative, qui sont utilisés pour produire du contenu à grande vitesse, peuvent être difficiles à prédire et sont beaucoup plus susceptibles de saper ce contrôle.*²³¹

Cet alignement politique sur la ligne idéologique du *Parti communiste chinois* constitue en effet un impératif pour le régime : « *Ces grands modèles doivent mettre en œuvre les valeurs fondamentales du socialisme* », a déclaré Rebecca Arcesati du

²³⁰ How China has 'throttled' its private sector (Financial Times 12 Sep 2024)
www.ft.com/content/1e9e7544-974c-4662-a901-d30c4ab56eb7

²³¹ China got a jump in the AI revolution by developing systems that could see and analyze the world with cutting-edge speed. The area of AI known as computer vision, which enables tracking and surveillance, aligns with Chinese leader Xi Jinping's emphasis on political control.

Despite that early success, the country was caught flat-footed by the public debut of OpenAI's ChatGPT in late 2022 and the generative AI craze it unleashed. Generative AI's large language models, which are used to produce content at speed, can be difficult to predict and are much more likely to undermine that control.

China Puts Power of State Behind AI — and Risks Strangling It. Liza Lin (Wall Street Journal, July 16, 2024)
www.wsj.com/tech/china-puts-power-of-state-behind-ai-and-risks-strangling-it-f045e11d

Mercator Institute for China Studies. Les développeurs d'IA générative doivent désormais faire face au défi que constitue l'alignement politique²³². »

D. De nouvelles stratégies de politique industrielle pour l'IA

Lors de la survenue de la pandémie de Covid-19, même la très influente revue du *MIT* reconnaissait l'échec de la Silicon Valley à créer des innovations utiles aux citoyens. Et cela est en partie lié au modèle économique publicitaire de ces sociétés dont l'objectif est de recueillir (ou extraire) le plus de données sur les individus, le plus rapidement possible afin de qualifier précisément des cibles publicitaires via des technologies de microciblage de plus en plus souvent associées à des IA :

La pandémie a mis en évidence un problème lancinant : les États-Unis ne sont plus très doués pour trouver les nouvelles idées et technologies en rapport avec nos besoins les plus fondamentaux. Nous sommes très doués pour concevoir des produits brillants, principalement des logiciels, qui rendent nos vies plus pratiques à bien des égards. Mais nous sommes beaucoup moins doués pour réinventer les soins de santé, repenser l'éducation, rendre la production et la distribution alimentaires plus efficaces et, d'une manière générale, mettre notre savoir-faire technique au service des plus grands secteurs de l'économie²³³.

Parallèlement aux dispositifs de financement des technologies existants (ou de subventions accordées dans le cadre d'appels à projets technologiques), des dispositifs innovants de financement des innovations technologiques devront aussi être mis en place. Ces dispositifs auront pour but d'aider au développement

²³² “These large models need to implement core socialist values,” Arcesati said. “There’s this challenge of political alignment that generative AI developers need to come to terms with.”

Chat Xi PT ? China’s Chatbot Makes Sure It’s a Good Comrade (Wall Street Journal, May 24, 2024)

www.wsj.com/tech/ai/chat-xi-pt-chinas-chatbot-makes-sure-its-a-good-comrade-bdcf575c

²³³ Covid-19 has blown apart the myth of Silicon Valley innovation (MIT Tech Review 25 Apr 2020)

www.technologyreview.com/2020/04/25/1000563/covid-19-has-killed-the-myth-of-silicon-valley-innovation/

de technologies d'intérêt général dans des secteurs non couverts par les financements actuels ou en raison de délais de réalisation difficilement compatibles. Il conviendra également d'élaborer des nouvelles mesures d'accompagnement pour les technologies de l'IA ayant le plus fort impact sociétal. Ainsi, comme le préconisent les experts de la *Rand Corporation* :

Les données et la puissance de calcul sont [ou devront être] largement accessibles aux entreprises, grandes ou petites, et aucune entité ne peut prédire avec certitude d'où viendra la prochaine avancée révolutionnaire en matière d'IA. Par conséquent, les gouvernements devraient donc élargir leurs outils au-delà des techniques de régulation traditionnelles. Deux mécanismes créatifs pourraient être mis en place : investir dans la création de bases de données publiques robustes pour la recherche sur l'IA ou accorder des subventions pour des défis visant des applications socialement bénéfiques de l'IA²³⁴.

Des initiatives privées ont aussi été conçues pour développer le principe du financement de « *missions technologiques* ». C'est le cas de la *Fondation XPRIZE*, organisation à but non lucratif fondée en 1995, qui a par exemple été utilisée pour lancer le défi *IBM Watson AI Xprize* consacré à l'amélioration des équipes mixtes « humains-IA »²³⁵.

Des plateformes de financement de défis technologiques spécifiques devront aussi être conçues pour aider à développer des technologies d'IA dans des domaines stratégiques. Cela a été le cas aux États-Unis avec la plateforme *Challenge.gov* créée en 2010 par l'administration Obama. À la différence du principe des appels à projets thématiques (comme ceux du programme *Horizon Europe*), cette

²³⁴ Data and computing power are widely available to companies large and small, and no single entity can reliably predict from where the next revolutionary AI advance might originate. Consequently, governments should consider expanding their toolboxes beyond traditional regulatory techniques. Two creative mechanisms could be for governments to invest in establishing robust, publicly owned data sets for AI research or issue challenge grants that encourage socially beneficial uses for AI. ‘

AI and Geopolitics : How Might AI Affect the Rise and Fall of Nations? (Barry Pavel, Ivana Ke, Michael Spirtas, James Ryseff, Lea Sabbag, Gregory Smith, Keller Scholl, Dominique Lumpkin (Rand Corporation, Nov 3, 2023) www.rand.org/pubs/perspectives/PEA3034-1.html

²³⁵ www.xprize.org/prizes/artificial-intelligence

plateforme avait pour objectif de financer des technologies répondant à des critères d'intérêt général sur la base de cahiers des charges précis établis par des collèges d'experts²³⁶.

Les 3 objectifs stratégiques d'une politique industrielle pour l'IA en Europe

- Réduire la dépendance vis-à-vis des technologies non-européennes en renforçant l'écosystème européen de l'IA via un important effort à l'échelle de l'UE, tout en mettant en place des régulations strictes en matière de concurrence, de transparence et d'interopérabilité des plateformes,
- Aider à développer des technologies d'IA éthiques qui éviteront de reproduire les dérives autocratiques ou liberticides des technologies extra-européennes tout en luttant contre les modèles économiques toxiques basés sur la captation massive de données personnelles,
- Développer l'écosystème européen pour les IA (et en particulier pour les IA stratégiques pour les échanges démocratiques et la gestion des données sensibles et stratégiques). Aider les entreprises européennes d'IA en leur facilitant l'accès à la commande publique et en élaborant des mécanismes de financement innovants.

²³⁶ Challenge.gov: Two Years and 200 Prizes Later (Obama White House Archive 5 Sep 2012)
<https://obamawhitehouse.archives.gov/blog/2012/09/05/challengegov-two-years-and-200-prizes-later>

XIV. RECOMMANDATIONS

Recommandation 1

Créer une vaste initiative de financement dédiée aux technologies de l'IA européenne via une procédure d'endettement commun de l'ensemble des pays de l'Union européenne. Cette initiative aura pour objectif d'aider au développement des entreprises, des filières clés et des technologies de l'IA en Europe. L'un des volets de cette initiative concernera le financement des infrastructures et des technologies nécessaires à la conception et au fonctionnement des IA en Europe.

Recommandation 2

Nommer un coordinateur des technologies de l'État. Créer à l'échelon français et européen la fonction de coordinateur des technologies de l'État, à l'instar du *CTO (Chief Technology Officer)* mis en place auprès de la *Maison-Blanche*. La mission de ce coordinateur des technologies, sera de promouvoir et mettre en œuvre les stratégies françaises et européennes des technologies développées ou utilisées par les administrations françaises et européennes, et de sensibiliser les acheteurs publics à la mise en œuvre des commandes publiques dans le domaine des technologies d'IA. Il coordonnera également une cellule de réflexion prospective sur l'impact du numérique sur l'autonomie stratégique européenne ainsi que sur les conséquences de ces technologies sur l'emploi et l'employabilité par branche professionnelle.

Recommandation 3

Créer un *Small Business Act* et une *Small Business Administration* en France et en Europe. Mettre en place un *Small Business Act* en France et en Europe pour réserver une part significative de l'ensemble des commandes publiques à des PME innovantes dans le domaine de l'IA, afin de leur permettre de se développer à partir de commandes solvables.

Recommandation 4

Créer un *European Buy Act* pour les entreprises technologiques de l'UE, en particulier pour les IA stratégiques et les systèmes de gestion des données sensibles. Il conviendra, dans le même temps, d'être en mesure de bloquer des rachats potentiellement hostiles dans le domaine de l'IA, en particulier pour les technologies qui assurent l'indépendance stratégique de l'Union européenne.

Recommandation 5

Développer les programmes nationaux et européens de mise à disposition de puissance de calcul et de stockage des données à coût réduit pour les structures de recherche et les PME innovantes dans le domaine de l'IA. Développer dans le même temps la mise à disposition des données protégées détenues par les institutions et les entreprises européennes via les espaces de données souverains (*data spaces*). Ces programmes devront ainsi avoir pour objectif de favoriser le développement d'architectures innovantes nécessaires à la conception et au fonctionnement des IA issues d'entreprises et institutions européennes.

Recommandation 6

Inclure les technologies d'IA dans la « taxe carbone » de l'UE. Inclure les technologies d'IA dans la révision prévue en 2026 du dispositif de « mécanisme d'ajustement carbone aux frontières » (MACF), aussi appelé « taxe carbone ». Il conviendra d'y intégrer les secteurs industriels liés aux technologies d'IA, afin de favoriser l'évolution de l'ensemble des pays partenaires de l'UE vers les technologies d'IA « à faible émission carbone », en fonction des énergies et des matières premières nécessaires à leur fonctionnement (pour les microprocesseurs et les infrastructures de cloud).

Recommandation 7

Développer des enseignements sur l'éthique de l'IA au sein des institutions scolaires et universitaires, ainsi que dans le cadre de la formation continue des agents de l'État. Créer des mesures de sensibilisation aux aspects éthiques de l'IA auprès des élèves et des étudiants. Mettre en place des concours pour la réalisation de projets d'intérêt général liés à la sensibilisation à la désinformation et aux dérives politiques de l'IA.

Recommandation 8

Développer des programmes de formation et d'accompagnement sur les savoir-faire de l'État en matière d'IA pour ses métiers traditionnels, et en particulier pour ses fonctions régaliennes, ainsi que sur l'impact de l'IA sur la démocratie et sur l'économie.

Recommandation 9

Favoriser le développement des architectures d'IA décentralisées « résidentes », et développer les programmes de recherche européens sur le fonctionnement décentralisé des intelligences artificielles européennes. Développer dans le même temps des programmes européens de R & D sur l'apprentissage fédéré pour l'IA et le chiffrement homomorphique pour la sécurité des données utilisées par des IA sur des structures distantes. Mettre en place un programme public de l'État pour le développement des technologies nécessaires au stockage et au traitement de long ou très long terme des données publiques.

Recommandation 10

Concevoir des dispositifs de financement innovants pour les projets d'IA d'intérêt général en Europe. Concevoir en France et en Europe des dispositifs de financement innovants qui associeront les acteurs publics et les industriels européens afin de développer des IA d'intérêt général dans les domaines de la santé, de la maîtrise de l'énergie et des transports. Ces dispositifs seront basés sur le principe des « défis technologiques » et devront permettre d'accélérer le développement des technologies qui ne seraient pas disponibles par la seule intervention des acteurs industriels existants.

XV. PRÉSENTATION DE L'ISN ET IDFRIGHTS

Ce rapport a été réalisé par *l'Institut de la Souveraineté Numérique* en partenariat avec l'institut *iDFRights*. Il a été dirigé par Bernard Benhamou, secrétaire général de l'ISN avec l'aide de Jean-Marie Cavada président et Colette Bouckaert secrétaire générale de l'institut *iDFRights*.

Bernard Benhamou est secrétaire général de l'ISN, il a exercé les fonctions de délégué interministériel aux usages de l'Internet auprès du ministère de la Recherche et du ministère de l'Économie numérique. Il a coordonné les premières conférences ministérielles européennes sur l'Internet des objets et l'inclusion numérique lors de la présidence française de l'Union européenne en 2008. Il a auparavant été le conseiller de la délégation française au Sommet des Nations unies sur la Société de l'Information. Il enseigne actuellement la gouvernance de l'Internet à l'université Paris I – Panthéon Sorbonne après avoir créé les premiers enseignements sur Internet et administrations à l'Éna et Sciences Po Paris.

L'INSTITUT DE LA SOUVERAINETE NUMERIQUE (ISN)

L'Institut de la Souveraineté Numérique (ISN) est une association loi 1901 à but non lucratif qui a pour mission de fédérer les acteurs du numérique et, au-delà, les acteurs économiques afin de créer une synergie sur les enjeux liés à la souveraineté numérique européenne. L'ISN s'est engagé depuis sa fondation en 2015 à faire la pédagogie et mobiliser les citoyens, et leurs représentants, sur les enjeux de la souveraineté numérique. L'ISN considère qu'il est nécessaire de promouvoir la protection de notre cyberspace, au même titre que celle de notre espace terrestre, maritime et aérien. L'ISN propose des actions et des mesures technologiques,

juridiques et politiques qui permettent de faire valoir la souveraineté numérique sur l'ensemble de nos ressources numériques et en particulier sur nos données. Les travaux de l'ISN auprès des interlocuteurs industriels européens, des institutions (collectivités, administrations, ministères, parlement) portent à la fois sur les thématiques traditionnelles de la souveraineté numérique ainsi que sur les perspectives d'évolution des technologies et leur impact sur la société. Qu'il s'agisse de la protection des données personnelles et industrielles, de la souveraineté sur les infrastructures numériques critiques ainsi que les transformations sociales, politiques et culturelles engendrées par les évolutions technologiques. L'objectif de l'ISN est aussi de permettre aux acteurs technologiques européens de mieux intégrer les notions de souveraineté numérique dans le développement de leurs travaux pour en faire un avantage concurrentiel en termes de confiance et de sécurité pour leurs usagers. Enfin, l'ISN souhaite contribuer à la transformation numérique de l'État afin de permettre de garantir la protection de notre souveraineté et de préserver dans le même temps nos libertés individuelles et collectives.

iDFRIGHTS : INSTITUTE FOR DIGITAL FUNDAMENTAL RIGHTS

Depuis 20 ans, les technologies numériques ont connu un développement planétaire, pris une place centrale dans nos vies individuelles et collectives, comme dans la conduite des entreprises. Par la somme incalculable de données que les individus ou les acteurs de l'économie ont fournies, les humains ont alimenté une nouvelle strate capable de guider nos vies et les organisations. Cette couche de datas, nourries souvent à notre insu, a alimenté des géants mondiaux de l'industrie digitale : américains d'abord, chinois ensuite, en attendant que les Européens y prennent leur place. Leurs pratiques commerciales, dénoncées jusque devant les tribunaux ou des Parlements, font fi des lois de concurrence en vigueur dans les pays libres. Aucune personne sensée n'est contre le digital, ses possibilités de développement, ni contre l'intelligence artificielle, ni les perspectives quantiques.

Mais l'avenir dépendra d'un choix drastique : voulons-nous, au motif des services immenses que le numérique offre à l'humanité, que l'Homme continue à dominer les machines qu'il a inventées, ou acceptons-nous de devenir les sujets asservis en partie ou totalement à l'objet qu'il a construit. Voulez-vous devenir l'esclave de votre smartphone, ou son maître ? Un peu partout dans nos pays démocratiques, des organisations se créent pour protester contre les pratiques du capitalisme numérique, des universités y travaillent, des procès en tous genres appellent les tribunaux au chevet de nos libertés, des Parlements enquêtent et voudraient aboutir à des lois régulatrices, et même des gouvernements s'arment pour mettre de l'ordre et faire passer le numérique du dollar vert américain ou celui du digital rouge chinois, sous le fronton de nos démocraties minées par tant de comportements hors la loi. C'est pourquoi, après de longues années passées au Parlement européen à combattre pour réguler ce secteur (RGPD, Directive « Droits d'auteur » pour les créateurs, « Droits voisins » pour protéger la presse), nous avons créé cet Institut. Avec des juristes, universités internationales, spécialistes du numérique et personnalités morales issues de la société civile, nous travaillons à la pédagogie des risques, à l'élaboration de doctrines à partir de cas concrets, à la promotion de normes et de standards sécurisant les problèmes rencontrés par les entreprises et organisations, et à soutenir législateurs et gouvernants nationaux et de l'UE, avec un matériel juridique modernisé.

XVI. REMERCIEMENTS

L’Institut de la Souveraineté Numérique et iDFrights tiennent à remercier tout spécialement pour leur contribution à l’élaboration de ce rapport les personnalités suivantes :

Catherine Morin-Desailly

Sénatrice de la Seine-Maritime, membre de la commission des affaires européennes. Présidente de la commission spéciale sur le projet de loi sur la sécurité et la régulation de l'espace numérique

Philippe Latombe

Député de la Vendée. Secrétaire de la Commission des lois constitutionnelles, de la législation et de l'administration générale de la République

Luc Julia

Ingénieur et informaticien franco-américain, spécialisé dans l'intelligence artificielle. Il est l'un des concepteurs de l'assistant vocal *Siri*. Ancien vice-président de *Samsung* chargé de l'innovation, il rejoint *Renault* en 2021 en tant que directeur scientifique.

Michael Nelson

Directeur des technologies et des affaires internationales *Carnegie Endowment for International Peace*. Ancien conseiller pour les technologies d'Al Gore, vice-président des États-Unis

Désirée Milosevic

Co-présidente du *RIPE* (Réseaux IP Européens Network Coordination Centre) Cooperation Working Group. Former special Advisor to the Chair Advisory Group *United Nations - Internet Governance Forum (IGF)*

Olena Kushakovska

Présidente de *SAP Labs* France, Co-dirigeante du *Conseil Industriel pour la Recherche en Intelligence Artificielle (ICAIR)*. Diplômée de l'Université de Kiev en mathématiques appliquées et de *l'École Polytechnique*

Laurence Devillers

Professeure en IA et éthique à *Sorbonne Université* et chercheuse au *CNRS* dans le laboratoire interdisciplinaire des sciences du numérique (*LISN*) à Saclay. Ses recherches portent sur l'interaction humain-machine, la détection des émotions, le dialogue oral et la robotique socio-affective.

Eric Saund

Chercheur en sciences cognitives et en intelligence artificielle. Ancien directeur de recherche au *Xerox Palo Alto Research Center (PARC)*. Membre de *l'AAAS (American Association for the Advancement of Science)* et du *Congressional Science & Technology Policy Fellowship*.

XVII. ORGANISATIONS PARTENAIRES

L'Institut de la Souveraineté Numérique et l'Institut *iDFRights* tiennent à remercier les organisations suivantes qui ont permis la réalisation de ce rapport.

PARTENAIRE PLATINUM



Depuis 2010, *Clever Cloud* édite un service de *Platform as a Service*, basé en Europe. Son objectif est d'aider les équipes de développement à mettre en production applications et services numériques, sur une infrastructure fiable, scalable, sécurisée et avec une tarification transparente. *Clever Cloud*, est convaincu que l'hébergement industrialisé permettra aux entreprises de travailler plus vite, d'être plus agiles sur leur marché, de se concentrer sur leur valeur ajoutée et de ne plus s'inquiéter quant à leur technologie d'hébergement. *Clever Cloud* milite pour la souveraineté des données, afin d'assurer à tous un espace numérique libre et autonome. *Clever Cloud* crée et contribue à de nombreux projets Open Source européens et privilégie des partenaires technologiques français et européens. *Clever Cloud* s'engage ainsi dans un écosystème de partenaires œuvrant chaque jour à la défense et à l'amélioration de notre souveraineté numérique européenne, notamment à travers son engagement à la présidence de *l'Open Internet Project*.

PARTENAIRE GOLD



Fondé en 1996, *Claranet* compte 3 500 collaborateurs répartis dans le monde entier et opère des services managés pour plus de 10 000 clients de tous secteurs d'activité. *Claranet* fournit des services de réseau, d'hébergement et d'applications gérées au Royaume-Uni, en France, en Allemagne, aux Pays-Bas, au Portugal, en Espagne, en Italie et au Brésil. *Claranet* compte aujourd'hui la moitié des entreprises du CAC40 parmi ses clients. *Claranet* accompagne la mutation numérique avec l'avènement de technologies et d'approches qui induisent des changements en profondeur : clouds publics, Big Data, DevOps, Agile, Machine Learning, Intelligence Artificielle, Blockchain, micro-services, modern workplace. Cette mutation crée autant d'opportunités à saisir (nouveaux modèles économiques, nouveaux usages dématérialisés, nouveaux marchés, réduction du « time-to-market ») que de complexité et de menaces (cyber-attaques, confidentialité des données critiques, régulations renforcées, nouveaux entrants). Plus qu'un fournisseur de services à qui on confie une mission opérationnelle définie, *Claranet* est un partenaire, qui guide, conseille et accompagne ses clients dans la stratégie et la mise en œuvre de ces transformations.

PARTENAIRES SILVER



La maîtrise par *Quicktext* de l'IA et du big data dans l'hôtellerie permet une offre unique pour structurer et partager les données des hôtels, interagir automatiquement avec les clients (grâce à son assistante virtuelle *Velma*), optimiser les ventes, le contenu, le marketing, alimenter la veille économique et stratégique. *Quicktext* est régulièrement classé première dans le domaine de l'IA dans l'industrie hôtelière par Hotel Tech Report. L'avance technologique de *Quicktext* a été reconnue, notamment lors des Global Tech Award à Austin, TX (USA), Premium Travel Awards à Shenzhen (Chine) et des Rencontres Internationales du Tourisme à Paris (France). Les plus grands groupes hôteliers internationaux ont ainsi fait confiance à *Quicktext* : Accor (FRA), Emaar (UAE), Hyatt (USA), Highgate (USA), Warwick (USA), Eurostars (SPA), H-Hotels (GER), Jaz Hotels (EGY) Palladium (SPA), Precise (DEU), Santa Fe (MEX).



BeTomorrow agence globale d'applications & de transformations numériques a réalisé des produits aujourd'hui utilisés par plusieurs millions d'utilisateurs. Depuis plus de 20 ans, nos équipes multidisciplinaires façonnent des solutions intelligentes avec rigueur et pragmatisme. Des start-ups aux grands groupes, *BeTomorrow* a accompagné plusieurs centaines de clients dans la conception, le développement et le déploiement de solutions : applications mobiles, web, cloud/data & IA /IoT & 3D. De l'audit à la stratégie au lancement sur le marché, du design au développement en passant par la formation et le coaching, à chaque

phase de leur évolution nous adaptons notre accompagnement aux besoins spécifiques de nos clients. Nos trois grands domaines d'expertise sont le développement produit (*digital factory*), la transformation numérique (*digital transformation*) ainsi que la R&D et l'innovation (*digital innovation*). Notre quête constante d'innovation et d'expertise technologique définit notre identité, et notre mode opératoire agile et unique nous permet de piloter les virages numériques les plus complexes.



Prometheus-X est une association dédiée au développement d'une infrastructure souveraine pour l'échange de données en Europe, dans le cadre de la stratégie européenne pour les données. Elle fédère plus de 500 organisations à travers l'UE pour la création de data spaces. Grâce à l'élaboration de data spaces souverains, *Prometheus-X* permet aux organisations et aux individus de partager et d'accéder aux données de manière sécurisée, interopérable et conforme aux réglementations européennes, notamment en ce qui concerne la protection des données et la souveraineté numérique. L'approche de *Prometheus-X* repose sur la création de communs numériques, c'est-à-dire des briques technologiques open source permettant à toute entité de créer et d'exploiter un data space sans être dépendante d'un acteur unique. Ces communs garantissent une infrastructure ouverte et éthique pour le partage de données dans divers secteurs, tels que l'éducation, les médias, la santé, le secteur juridique et le tourisme. En consolidant un écosystème européen d'échange de données souverain, éthique et sécurisé, *Prometheus-X* souhaite jouer un rôle essentiel dans la garantie de l'indépendance technologique de l'Europe.



**EXPLORER ET DÉVELOPPER
LES TERRES INCONNUES**

Fondée en 1993, *ALCIMED* est une société de conseil spécialisée dans l'innovation et le développement de nouveaux marchés. Elle intervient principalement dans les secteurs des sciences de la vie (santé, biotechnologies, agroalimentaire), de la chimie, de l'énergie et de l'environnement, de la cosmétique et du luxe, ainsi que dans les domaines de l'aéronautique, du spatial, de la défense et des politiques publiques.

Grâce à ses collaborateurs hautement qualifiés, *ALCIMED* accompagne les grands groupes industriels, les entreprises de taille intermédiaire (ETI) et les acteurs institutionnels dans l'exploration et la conquête de leurs « terres inconnues » : technologies émergentes – notamment l'intelligence artificielle (IA) et ses applications sectorielles –, développement de nouvelles offres, expansion géographique, prospective et nouveaux modèles d'innovation. L'intelligence artificielle constitue un axe stratégique majeur pour *ALCIMED* : elle représente à la fois un territoire à explorer, une compétence à intégrer, un outil d'analyse au service de nos missions, et une opportunité de réaffirmer nos engagements en matière de souveraineté, d'éthique et d'excellence. *ALCIMED* est implantée en France (Paris, Lyon, Toulouse), ainsi qu'en Allemagne, en Italie, aux États-Unis et à Singapour.



INSTITUT DE LA
SOUVERAINETÉ
NUMÉRIQUE



Ce rapport a été réalisé par l'Institut de la Souveraineté Numérique (ISN) en partenariat avec iDFRights et il a été dirigé par Bernard Benhamou, secrétaire général de l'ISN (bernard.benhamou@souverainetenumérique.fr). Sauf mention explicite, les traductions qui figurent dans ce rapport ont été effectuées par les auteurs. Le texte de ce rapport (à l'exception des illustrations) est disponible sous licence Creative Commons : Attribution - Pas d'utilisation commerciale - Partage dans les mêmes conditions 4.0 France (CC BY-NC-SA 4.0 FR) © 2026 Institut de la Souveraineté Numérique (ISN) et iDFRights

