

An aerial photograph of a river delta, showing a complex network of channels and distributaries. The image is overlaid with a white grid. The text is printed in large, bold, white, sans-serif capital letters.

**INTELLIGENCE
ARTIFICIELLE :**

**POUR
L'ÉMERGENCE
D'ALTERNATIVES**

**REFONDER UNE
POLITIQUE
SCIENTIFIQUE,
INDUSTRIELLE ET
SOCIALE DE L'IA**

X - ALTERNATIVE

Introduction : une nouvelle alliance avec la machine

« *Pouvons-nous imaginer des technologies de l'information et de la communication qui ne nous exploitent, ne nous trompent et ne nous supplantent pas ? Oui, nous le pouvons - une fois que l'on sort des réseaux de pouvoirs commerciaux qui ont défini la vague actuelle de l'IA* », affirme l'écrivain britannique James Bridle dans un article du Guardian de 2023¹.

Portés par des acteurs privés monopolistiques aux objectifs clairement affichés de contrôle et de monétisation des comportements individuels², ces « *réseaux de pouvoir* » orientent le développement de l'Intelligence Artificielle vers de sombres destinées. Prolétarianisation des individus, déséquilibres économiques extrêmes, aliénation des travailleurs et travailleuses, impacts énergétiques et environnementaux démesurés : l'industrie de l'IA actuelle concentre et exacerbe toutes ces « externalités négatives ». C'est à se demander si son développement actuel ne serait pas l'ultime parade du capitalisme néo-libéral.

Est-ce pour autant tout ce qu'il y aurait à dire sur un champ de recherche scientifique aussi fascinant ? N'y aurait-il aucun futur souhaitable, alternatif, pour les technologies de l'intelligence artificielle ? Souhaiter des alternatives ne suffit pas, il faut s'atteler à les faire apparaître : se demander à quoi elles pourraient ressembler, et à quelles conditions elles pourraient émerger.

Nous nous proposons ici de faire un pas de côté, d'adopter une réflexivité sur la technique instruite de ses rapports aux conditions socio-économiques, aux problématiques éthiques comme pratiques, aux jeux d'influences et aux discours médiatiques des garants du techno-solutionnisme. Repenser l'Intelligence Artificielle non pas comme une entité semi-consciente annonciatrice d'une ère nouvelle, mais comme un mode d'assemblage de méthodes statistiques, de jeux de données, de puissance de calcul, et d'idéologies. Ancrer notre critique de l'IA dans le réel en parcourant son histoire, en investiguant ses modes spécifiques de développement scientifique et industriel, en questionnant les narratifs en vogue et les intérêts qui les sous-tendent.

Souligner enfin ce sur quoi il faudrait agir pour soutenir l'émergence, en France et en Europe, de voies alternatives pour ces technologies. S'il s'agit, comme nous le soutenons, de mettre les technologies numériques au service des décisions collectives, c'est-à-dire aussi de nos capacités d'invention, d'imagination et d'interprétation, il nous faut porter une vision alternative de "l'intelligence". Selon l'association Ars Industrialis, « ce qui est bête ou intelligent, ce n'est pas tant tel individu ou tel milieu que la relation qui les lie l'un à l'autre ». L'enjeu pour nous n'est donc pas de déterminer quelles "solutions" techniques feraient progresser l'IA, mais plutôt de construire et décrire une « nouvelle alliance avec la machine » qu'appelait de ses vœux en 1992 le philosophe Félix Guattari³.

¹ [The stupidity of AI](#), James Bridle, The Guardian, 16/03/2023 (traduction des auteurs)

² A ce sujet, nous ne saurions trop insister sur l'importance de l'éclairage de cet ouvrage : *L'Age du capitalisme de surveillance*, Shoshana Zuboff, Editions Zulma (pour la traduction française), 2018

³ [Pour une refondation des pratiques sociales](#), par Félix Guattari (Le Monde diplomatique, octobre 1992)

X-Alternative est un groupement d'ingénieurs et scientifiques au service de l'intérêt général : en ouvrant des espaces de réflexion, nous tâchons de repenser la technique pour une inscription sociale émancipatrice. Composé de jeunes chercheuses en informatique, d'industriels expérimentés, de philosophes des techniques, le comité de rédaction de cette note s'est donné pour tâche de poser un regard technique et critique sur l'Intelligence Artificielle et les problématiques qu'elle soulève. Cette note s'adresse aux décideurs, fonctionnaires, cadres, chercheurs, étudiants, au grand public : à qui-conque chercherait un partage sensible entre mirages de quelques-uns et réalité de tous.

Nous souhaitons ainsi proposer le socle sur lequel refonder une politique scientifique, industrielle et sociale de l'Intelligence Artificielle.

Cette note se structure en trois moments. Nous partons du *constat que l'Intelligence Artificielle est un projet technoscientifique empreint d'idéologies*. En retraçant brièvement la généalogie d'un champ scientifique et industriel, nous tenterons de désigner les embranchements, les voies privilégiées, qui laissent imaginer tout ce que cette technologie aurait pû - et pourrait encore - être. Nous nous joindrons ensuite aux nombreuses voix qui dénoncent les conséquences psychologiques, sociales, économiques, environnementales que le mode de développement actuel de l'Intelligence Artificielle provoque. Nous évoquerons enfin quelques exemples des récupérations politiques auxquelles les imaginaires qui imprègnent ce champ technologique laissent libre cours.

Nous soulignons ensuite les *points névralgiques des structures de développement et de contrôle de l'IA*. Revenir brièvement sur le fonctionnement d'une intelligence artificielle nous permettra de dévoiler les conceptions fondamentalement politiques qui se logent dans leurs rouages. Nous nous attacherons ensuite à décrire le fonctionnement actuel de la recherche fondamentale et industrielle en IA et ses implications sur le développement de ces technologies. Enfin, des considérations industrielles, législatives et géopolitiques achèveront de rendre compte des points stratégiques qu'une politique industrielle devrait adresser.

Cet enchaînement nous conduira à formuler des *propositions pour une politique industrielle des automates numériques computationnels*. Ces propositions s'alignent sur notre engagement plus large pour un repositionnement stratégique de nos capacités industrielles, et plus particulièrement pour un intérêt retrouvé de la part des puissances publiques vis-à-vis de notre souveraineté numérique.

X-Alternative, Groupe de travail Intelligence Artificielle

Rapport final :

Intelligence Artificielle : pour l'émergence d'alternatives
Refonder une politique scientifique, industrielle et sociale de l'IA
Novembre 2024

*image de couverture : vue satellite de la mine de Chuquicamata, Chili (22°17'20"S
68°54'07"W 2 258 m), @Airbus via Google Earth, 2023*

Ce rapport est l'aboutissement d'un travail de réflexion collaboratif porté par l'association X-Alternative.

Il n'aurait pu être possible sans les apports significatifs des membres du groupe de travail : Zako Sapey-Triomphe, Samuel Laval, Nilo Schwencke, V., Victor Chaix, Pierre Musacchio, Auguste Lehuger, Lucas de Lara, Dominique Marre, Alexis Leautier

En plus de nombreux contributeurs extérieurs, que nous remercions chaleureusement.

Si vous souhaitez être au courant de nos actualités, inscrivez vous à notre [Newsletter d'information](#), et suivez X Alternative sur nos canaux de diffusion :

@X_Alternative (twitter, bluesky)

@XAlternative (youtube)

<https://x-alternative.org>

I n d e x

	Introduction : une nouvelle alliance avec la machine	2.
	Résumé exécutif	6.
<hr/>		
I.	I. Constats : L'Intelligence Artificielle, un projet technoscientifique empreint d'idéologies	
	I.1 Généalogie des automates numériques : de la cybernétique à «l'IA générative»	9.
	I.1.a Personnages clés	9.
	I.1.b Moments de pivot	11.
	I.2 L'IA des GAFAM aux quatre coins du monde : quels impacts politiques, économiques et sociaux ?	13.
	I.2.a Sur les individus	13.
	I.2.b Sur la concentration économique	14.
	I.2.c Sur le salariat	15.
	I.2.d Sur l'environnement	15.
	I.3 Un sésame politique	16.
	I.3.a Altruisme et accélérationisme efficaces	16.
	I.3.b La planification écologique	17.
	I.3.c Numérique et gouvernementalité	17.
	I.4 Possibilités d'émancipation et d'organisation collective	18.
<hr/>		
II.	II. Points névralgiques : les structures de développement et de contrôle de l'IA	
	II.1 Sur le fonctionnement d'une ia	20.
	II.1.a Termes importants et critiques associées	21.
	II.1.b Domaines de recherche et d'applications	22.
	II.2 La recherche scientifique	25.
	II.2.a La qualité de la recherche et ses évolutions récentes	25.
	II.2.b Les choix de carrières possibles pour un.e jeune chercheur.se	27.
	II.2.c Comparaison avec d'autres domaines	28.
	II.3 Considérations géopolitiques	29.
	II.3.a Le poids des acteurs américains :	29.
	II.3.b La place de la Chine	29.
	II.3.c. L'UE essaie d'exister :	30.
	II.4 Législation et contrôle de l'État	30.
<hr/>		
III.	III. Propositions : pour une politique industrielle des automates computationnels	
	III.1 Définir collectivement le projet politique qui doit orienter le développement de l'IA	33.
	III.2 Prendre des mesures d'alignement et de restructuration de la recherche publique et privée	35.
	III.3 Engager l'ensemble de la chaîne industrielle	36.
	III.4 Aligner les financements publics	38.
	III.5 Poursuivre le mouvement de régulation pour favoriser une approche européenne du domaine	38.
	III.6 Proposer une alternative aux discours médiatiques actuels	39.
	III.7 Investir et promouvoir une conception démocratique des IA	39.
	III.8 Normaliser les pratiques industrielles pour minimiser les impacts néfastes et renforcer la cybersécurité.	40.
<hr/>		
	Conclusion	41.

Résumé exécutif

Entre promesses techno-solutionnistes, prophéties millénaristes ou peurs infondées, les récents développements de l'Intelligence Artificielle (IA) sont quotidiennement débattus et analysés, à tel point qu'il devient difficile de discerner ce que le terme même recouvre. Nous ne formulerons pas ici de définition positive de l'ensemble de sciences et technologies qui se retrouvent aujourd'hui qualifiées d'Intelligence Artificielle (qui appellerait par ailleurs une critique du terme lui-même) : nous parlerons de tout ce qui est couramment qualifié d'IA.

Ingénieurs, chercheuses, ou philosophes des sciences, nous souhaitons formuler une critique d'une part, et discerner d'autre part les causes qui font de cette techno-science ce qu'elle est devenue. Désigner enfin ce qu'il nous semble nécessaire pour faire émerger des alternatives à son mode de développement actuel, captif d'intérêts économiques aux conséquences néfastes.

Parmi d'autres, nous voyons quatre impacts négatifs de l'IA aujourd'hui :

1. **Sur les individus** : des systèmes dont l'unique fonction est de paraître humains dans le but de forcer plus loin le déploiement de dispositifs d'extraction de données portent en eux la prolétarianisation croissante des individus et la captation des attentions.
2. **Sur la concentration économique** : la position hégémonique des GAFAM et la structure de leurs opérations leur permettent une captation inédite de la valeur ajoutée, en plus du fruit de nos interactions quotidiennes sur leurs plateformes.
3. **Sur le salariat** : la violence sociale des chocs potentiels sur le marché de l'emploi se couple avec le vide intellectuel que l'automatisation des tâches intellectuelles imposerait au travail.
4. **Sur l'environnement** : l'impact environnemental calamiteux de technologies dont la "performance" dépend directement de leur intensité énergétique et dont les propriétaires n'ont pour toute réponse à apporter au changement climatique qu'un discours prophétique techno-solutionniste.

Et nous croyons fermement qu'il pourrait en être autrement, que ces systèmes portent en eux d'autres promesses, celles de soutenir de nouvelles capacités industrielles, d'inventer d'autres manières de faire démocratie, de repenser profondément notre rapport aux machines numériques. Alors que faire ?

Pour imaginer l'alternative, il nous faut d'abord retracer ce qu'il peut y avoir de contingent, d'accidentel, voire d'idéologique dans l'ensemble des technologies qui se retrouvent qualifiées d'intelligences artificielles. Dans les années 1950, Norbert Wiener, Franck Rosenblatt, et d'autres personnages, qui ne cachent pas leur engouement pour la théorie du marché de Friedrich Hayek ou pour les hypothèses cognitivistes sur le fonctionnement du cerveau, fondent ce champ scientifique. L'IA se trouve rapidement mêlée, dans son développement industriel, aux grands mouvements du capitalisme américain et de son complexe militaro-industriel. Une brève généalogie nous montre les chemins qui, du côté de la recherche soviétique ou peut-être en se faufilant hors du filet financier de la crise des dot-com, auraient pu mettre quelques billes sur des pentes différentes.

Continuer l'analyse : quels sont les points d'inflexion, les lieux stratégiques où nous pourrions porter nos efforts ?

Ouvrons d'abord la boîte noire, et voyons comment fonctionne une ia⁴. Comme une fonction entrée/sortie qui répond à une tâche précise, sur la base de paramètres définis lors

⁴ Sur la différence que nous faisons entre IA et ia, voir [l'Exergue 1 : IA ou ia ?](#), p. 10

d'un **entraînement** sur des **données**. Cet entraînement vise à optimiser une **fonction de coût** définie selon les **objectifs** que le système vise à remplir. Entrée/sortie, paramètres, entraînement, données, fonction de coût, objectifs : tout cela est défini par l'ingénieur qui conçoit le système. Les données peuvent avoir des biais, les couples entrée/sortie être difficiles à évaluer et encadrer, les paramètres sont peu interprétables et l'entraînement est une étape relativement empirique. Nous soutenons toutes ces critiques, nous en portons une dernière à l'attention de notre lecteur : quels sont les objectifs, et de quelle manière sont calculées les fonctions de coût ? Qui décide de ce que doit chercher à accomplir le système ? L'IA est fondamentalement une science de l'optimisation : que cherche-t-on à optimiser ?

Optimisez le temps passé sur un écran, vous obtiendrez les réseaux sociaux polarisateurs et divergents⁵. Optimisez la prédiction des fraudes aux allocations sociales, vous obtenez un système multiplicateur de discriminations⁶. Optimisez la vraisemblance statistique de fragments de mots, vous obtenez ChatGPT. Il y a un lien direct entre les décisions scientifiques et techniques et l'usage final, les conséquences que ces systèmes portent sur leurs utilisateurs.

Ces biais, ces objectifs, ces rouages choisis parfois arbitrairement par une poignée de scientifiques et d'ingénieurs, il nous faut les reconquérir politiquement. Comment se structure le champ de développement et de contrôle de l'IA ?

La recherche scientifique en IA se distingue des autres domaines de recherche à la fois par son fonctionnement et par les acteurs qui la dominent. Le parallèle avec l'industrie du tabac ou la recherche pharmaceutique n'est pas loin : l'intelligence artificielle demeure cet agrégat étrange entre domaine de recherche scientifique, ensemble de technologies, et marché en plein essor, tous trois dominés par une poignée d'acteurs dont les capacités financières et l'expertise en politique industrielle talonnent celles de certains pays du G20. Le raccourcissement du processus d'innovation, de la recherche fondamentale à la mise sur le marché en quelques années voire quelques mois, prend ouvertement ses sources dans l'accélérationisme⁷. Les exigences de rentabilité à court terme des marchés et la faiblesse des freins réglementaires renforcent ce mouvement. Des pans entiers de la production scientifique s'alignent sur ces impératifs, comme en témoigne l'influence qu'ils exercent sur les principales conférences du domaine. Réorientés vers la recherche et le développement, des laboratoires privés aux moyens colossaux peuvent d'autant plus facilement donner le « la » dans ces domaines où les infrastructures de calcul coûtent cher et où les plus offrants recrutent les têtes les mieux formées.

Et cela a de graves conséquences sur la qualité de la recherche et les perspectives qui s'ouvrent aux chercheurs.ses et ingénieurs.es en début de carrière.

Mais la recherche s'inscrit dans un contexte industriel et social, et il faut souligner les intrications notoires entre la recherche en IA et les complexes militaro-industriels des pays qui la soutiennent. Sans approfondir une analyse de relations internationales qui nous porterait loin de notre champ de compétences, nous remarquons la place apparemment cruciale que joue l'IA dans les conflictualités entre Chine et Etats-Unis, ainsi que le profond déséquilibre infrastructurel que révèle le contrôle quasi-monopolistique d'une seule entreprise (NVIDIA) sur le marché des processeurs.

⁵ voir à ce sujet le travail remarquable de David Chavalarias, dans *Toxic Data*, Flammarion Essais, 2022

⁶ <https://www.laquadrature.net/2023/11/27/notation-des-allocataires-lindence-des-pratiques-de-la-caf-desor-mais-indeniable/>

⁷ Pour une description de l'accélérationisme (*Effective Accelerationism*), ses racines et ses influences, voir [I.3.a Altruisme et accélérationisme efficaces](#), p. 16

Enfin, l'appareil législatif de l'Union Européenne, dont les éloges font grand bruit, et bien qu'il soit de fait plus avancé que bien d'autres tentatives en la matière, paraît servir à l'instauration d'un état de fait et semble encore loin de se montrer sourd aux lobbying de puissances étrangères.

Ces considérations multiples nous amènent à formuler nos futurs souhaitables, autour de huit propositions :

1. **Définir collectivement le projet politique qui doit orienter le développement de l'IA.** Animer d'abord un débat public sain et informé sur une question rarement posée alors qu'elle devrait être première : l'IA, pour quoi faire ? A travers le concept de communs numériques, nous proposons de faire des automates des instruments de la puissance publique.
2. **Prendre des mesures d'alignement et de restructuration de la recherche publique et privée.** Aider les jeunes chercheurs.es et ingénieurs.es à s'orienter. Multiplier les structures de recherche indépendantes, mutualiser les capacités de calcul, et redéfinir les liens avec l'industrie et la commande publique.
3. **Décliner les mesures d'alignement sur l'ensemble de la chaîne industrielle.** En passant par une reprise en main des politiques industrielles par l'État, porter une approche intégrée qui vise à encourager et encadrer l'IA pour l'industrie, pas pour les consommateurs.
4. **Aligner les financements publics.** A la fois à travers les plans d'investissement (comme France 2030), mais aussi le crédit impôt-recherche et les aides publiques à l'innovation.
5. **Poursuivre le mouvement de régulation pour favoriser une approche européenne.** Se servir ouvertement de la loi pour pousser nos propres acteurs et notre vision : renforcer les obligations de transparence, imposer la possibilité d'utiliser des alternatives aux algorithmes dominants, envisager un nouveau cadre de propriété intellectuelle.
6. **Regagner du terrain médiatique en donnant à voir des alternatives.** Rendre concret des possibilités de développement alternatives pour sortir des discours fatidiques. Fournir des éléments de contexte pour réaffirmer le caractère politique et économique de la technologie.
7. **Investir et promouvoir une conception démocratique de l'IA.** Financer des initiatives collaboratives pour développer des systèmes sur lesquels le contrôle démocratique est assuré "by design".
8. **Normaliser les pratiques industrielles pour minimiser les impacts néfastes et renforcer la cybersécurité.** Instaurer un principe de non-conformité pour assurer les standards les plus stricts, adossés à une capacité d'audit et de contrôle européenne.

Nous demeurons convaincus que la mise en œuvre intelligente et ambitieuse des principes formulés dans cette note nous rendrait le contrôle des politiques industrielles, scientifiques et sociales de l'intelligence artificielle, nous permettant de la remettre au service de l'intérêt général et d'ouvrir son à venir.

I. Constats : L'Intelligence Artificielle, un projet technoscientifique empreint d'idéologies

*Attardons nous dans un premier temps sur quelques constats, qui nous permettront de poser les repères pour situer notre critique, et d'informer notre lecteur sur ce qui nous semble aujourd'hui problématique dans le projet industriel et scientifique de l'Intelligence Artificielle. Nous reprenons à dessein ce terme, nullement pour occulter les nombreuses critiques qui peuvent lui être faites - et nous rejoignons d'ailleurs d'autres voix qui voudraient lui substituer le terme plus adapté «d'automate computationnel» - mais pour embrasser sciemment les multiples destinations et les nombreuses ambiguïtés qui s'y logent. Nous formulons une définition concise et volontairement flexible des termes «IA» et «ia», voir *Exergue 1 : IA ou ia ?*, p. 10.*

I.1 Généalogie des automates numériques : de la cybernétique à «l'IA générative»

Commençons par le commencement : les machines apprenantes, la cybernétique, l'intelligence artificielle.

I.1.a Personnages clés

Origines (années 1950)

Dès 1950, le célèbre mathématicien et précurseur des sciences informatiques Alan Turing, dans son article «*Computing machinery and intelligence*»⁹, annonce vouloir répondre à la question : “est-ce que les machines peuvent penser ?”, et pose alors une référence fondamentale, encore utilisée aujourd'hui. À la question des machines pensantes, Turing substitue en réalité son jeu de l'imitation, dans lequel un opérateur doit tenter de déterminer le sexe (homme/femme) de ses deux interlocuteurs, dont l'un a pour objectif de tromper l'examineur. Selon Turing, si, en remplaçant un des humains par une machine, le taux d'erreur de l'examineur ne variait pas, nous aurions alors une machine capable de penser. C'est donc en imaginant une machine ayant pour fonction d'imiter le mensonge que Turing lance la quête de l'intelligence artificielle.

En 1956, la ville américaine de Dartmouth accueille un groupe de mathématiciens pour le Dartmouth Summer Research Project on Artificial Intelligence, un séminaire de recherche qui établira l'usage du terme pour désigner des systèmes simulant l'esprit humain. John McCarthy impose la formule pour se démarquer de Norbert Wiener et des cybernéticiens qui, à l'époque, concentrent l'attention et les ressources attribuées à l'automatisation de processus industriels¹⁰. Ce nouveau domaine d'étude postulait d'emblée que « l'esprit était quelque chose d'ordonné, qu'il vivait à l'intérieur du cerveau individuel et qu'il suivait une logique implicite et fiable, qui pourrait être modélisée de manière convaincante par des modes computationnels dérivés de l'observation d'évènements sociaux »¹¹.

⁸ Anne Alombert et Giuseppe Longo, «[Il n'y a pas d'intelligence artificielle : parlons d'automates numériques pour rompre avec les idéologies publicitaires !](#)», *L'Humanité*, Saint-Denis, 11 juillet 2023.

⁹ Turing, A.M. (1950). *Computing machinery and intelligence*. *Mind*, 59, 433-460.

¹⁰ David J Gunkel, « [What's in a Name? Cybernetics vs AI](#) », *sublationmedia.com*, 19 juin 2023.

¹¹ Jonnie Penn, « *Animo nullius: on AI's origin story and a data colonial doctrine of discovery* », *BJHS Themes*, Cambridge, Vol. 8, 2023.

Exergue 1 : IA ou ia ?

Nous ne chercherons pas dans cette note à poser de définition positive et exhaustive de l'Intelligence Artificielle, mais nous devons cependant poser d'emblée deux acronymes qui nous seront utiles :

Par **Intelligence Artificielle (IA)**, nous entendons un champ de recherche scientifique et une chaîne industrielle complexe, regroupant un ensemble de technologies disparates et parfois fondamentalement différentes, qui visent généralement à automatiser des tâches habituellement attribuées à l'entendement humain (mais pas uniquement).

Une **intelligence artificielle (ia)** désigne un exemple de ces technologies, comme un système de traduction (DeepL), ou de transcription vocale (Whisper), ou de génération d'images (MidJourney).

Ainsi nous dirons : « la consommation énergétique de l'IA est équivalente à celle d'un pays du G20 » et « une ia comme ChatGPT a d'abord pour fonction d'imiter l'expression écrite d'un humain ».

Inspirations des théories économiques libérales

Extrapoler les comportements humains à partir d'un modèle théorique d'individu rationnel et calculateur : d'emblée les méthodes d'IA s'inspirent de l'économie orthodoxe. Herbert Simon, l'un des pionniers de ce domaine et lui-même économiste, puisera dans les études d'Adam Smith sur l'administration et les processus de prise de décision pour orienter ce qui fondera le « *paradigme symbolique* » de l'intelligence artificielle : la conception de systèmes associant des séries de règles de décision conçues par des spécialistes. Le psychologue Franck Rosenblatt trouvera dans les travaux de Friedrich Hayek sur les structures de marchés, associations décentralisées et spontanées, l'inspiration pour son perceptron, ancêtre des « *réseaux de neurones* » et emblème du « *paradigme connexionniste* » : selon ce modèle, l'intelligence artificielle doit faire émerger un ordre naturel capable d'organiser statistiquement le monde de manière plus efficace, fonctionnelle, rationnelle que les individus et les instances collectives telles que des États.

Ces deux domaines, paradigme symbolique et paradigme connexionniste, souvent opposés au sein des sciences de la computation (Computer Science), découlent en réalité des mêmes présupposés. Comme le défend le philosophe Mathieu Triclot, c'est la notion floue « *d'information* » qui a permis des analogies entre des choses aussi éloi-

gnées que les machines et le vivant, les calculateurs et le cerveau, l'économie politique et la métaphysique. Ces analogies ont marqué tout autant les deux paradigmes de l'IA : symbolique et connexionniste. De tels parallèles contredisent d'ailleurs le discours des cybernéticiens et précurseurs de l'informatique, comme Von Neumann, pour qui « *le traitement de l'information qui a lieu dans le cerveau est profondément différent de ce qui se passe dans un ordinateur* »¹².

Figures modernes

La recherche scientifique en IA s'incarne aujourd'hui dans les figures de MM. Joshua Bengio et Yann le Cun, tous deux récipiendaires du prix Turing en 2018. Le premier, professeur d'informatique à Montréal, prône la régulation et insiste sur les dangers du développement sans frein de l'IA. Le second dirige la recherche en intelligence artificielle chez Meta (ex Facebook) dont il est aussi vice-président, et se veut plus rassurant – ayant peut-être à cœur les intérêts de son groupe. Si leurs discours les opposent souvent, tous deux portent un regard calculatoire et individualisant sur l'intelligence humaine, dans la lignée directe de McCarthy et Rosenblatt.

I.1.b Moments de pivot

Hiver de l'IA (années 80)

À la fin des années 1980, la recherche dans ce domaine entame un long hiver. Le terme même d'intelligence artificielle décourage les financeurs et l'IA prend le nom « *d'algorithmie avancée* ». En URSS, les scientifiques développent une autre approche : formés par d'autres présupposés idéologiques que leurs confrères occidentaux, les chercheurs soviétiques repoussent l'idée d'une machine pensante imitant la rationalité d'un acteur individuel. L'intelligence humaine, estiment-ils, émerge comme une faculté sociale et culturelle dans la multitude des interactions entre individus. McCarthy et ses collègues inscrivaient l'IA dans l'idéologie du libre marché et les sciences cognitives en tentant de répliquer les cognitions individuelles ; l'IA "communiste" cherche au contraire à répliquer l'intelligence émergeant des macrostructures trop complexes pour l'entendement humain. Leur projet s'institutionnalise en 1989 avec la fondation de l'Association Soviétique de l'IA, au moment précis où le système soviétique vacille¹³. Mais si le projet soviétique avait pu voir le jour, les deux approches se seraient peut-être retrouvées en fin de course. En effet, pour mettre au point l'IA américaine et connexionniste, le capitalisme a dû collecter massivement des données, et pour imaginer l'IA communiste, les soviétiques reconnaissaient qu'il faudrait « *assimiler la culture et le sens humain* »¹⁴. Quoi qu'il en soit, lorsque la recherche en IA émerge à nouveau, au début des années 90, elle le fait dans un monde profondément changé.

¹² Mathieu Triclot, « La notion d'information dans la cybernétique », journée « histoire et didactique des sciences », Lyon, 6 décembre 2004.

¹³ Olessia Kirtchik, « The Soviet scientific programme on AI: if a machine cannot 'think', can it 'control'? », BJHS Themes, Cambridge, Vol°8, 2023.

¹⁴ *ibid*

Boom technologique (années 90-2000)

Aux États-Unis, dès l'année 1990, Internet devient le pinacle du monde technologique et les investisseurs surexcités financent à nouveau sans retenue les départements de recherche concernés par ce domaine. Du moins jusqu'à l'explosion de la bulle Internet en 2000. Désormais soumises aux pressions de leurs bailleurs de fonds, les entreprises numériques engagent une course à la rentabilité, laquelle passe par la publicité. Google met au point la marchandisation des données personnelles captées en masse chez les utilisateurs de services gratuits et utilisées comme matière première pour prédire puis orienter les comportements.

La nouvelle vague de l'IA prend ainsi racine dans ce que Shoshana Zuboff qualifie de « *capitalisme de surveillance* »¹⁵. Elle se concrétise au début des années 2010 avec la publication du challenge annuel de reconnaissance d'images ImageNet, construit sur la base de données éponyme. Cette production titanesque s'inscrit dans un travail de longue haleine initié dès 2006 par la chercheuse Fei Fei Li. Contrairement à nombre de ses collègues, cette dernière se concentre non pas sur les algorithmes et la puissance de calcul, mais sur la donnée : une ressource essentielle à l'accomplissement de la prophétie connexionnisme et dont le développement d'un Internet sous contrainte commerciale permet l'extraction massive.

Ainsi, à partir de 2011, l'IA se trouve façonnée par ce technocapitalisme dont elle dépend. Les géants du web et leur armée d'annotateurs mobilisés par la plateforme Amazon Mechanical Turk aux quatre coins du globe, structurent et raffinent des bases de données. L'édition 2012 du challenge ImageNet voit le triomphe de « *l'apprentissage profond* », invention la plus ambitieuse des connexionnistes, jusqu'alors considérée comme irréalisable. Ses évolutions récentes, comme ChatGPT, ont nécessité des milliers de travailleurs kényans, payés moins de deux dollars l'heure pour vérifier manuellement les réponses produites par le système¹⁶.

Transformation d'OpenAI

Dans l'imaginaire collectif s'ancre l'idée que les progrès de l'IA dépendent de l'accroissement presque infini du nombre de paramètres et donc des capacités de calcul, qui entraîne tout l'appareil de production vers une fuite en avant. L'organisation à but non lucratif OpenAI est fondée en 2015 avec la participation de Peter Thiel et Elon Musk, figures du capitalisme débridé. Elle entend promouvoir et développer « *une intelligence artificielle générale sûre et d'en partager les bénéfices avec le monde entier* », comme l'indique sa charte d'honneur. Pour en assurer le respect, OpenAI se dote d'un conseil d'administration scientifique composé de figures reconnues dans la recherche en éthique sur les systèmes d'IA, comme Mme Helen Toner. Quatre ans plus tard, l'organisation se transforme en entreprise lucrative à profit plafonné – 100 fois l'investissement initial.

¹⁵ Shoshana Zuboff, « Un capitalisme de surveillance », Le Monde diplomatique, janvier 2019.

¹⁶ Billy Perrigo, Exclusive: *OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic*, in Time Magazine, 18/01/2023

Aux scientifiques humanistes des premières heures se mêlent des ingénieurs ambi-
tieux, attirés par les stock-options. En 2020, OpenAI lance GPT3 tout en maintenant de
grands projets « éthiques », comme le « *superalignement* » (alignement de systèmes d'IA
supérieures sur des objectifs et des principes définis par des humains), dirigé par Ilya
Sutskever et Jan Leike. En novembre 2023 apparaît un changement de cap majeur : au
terme d'une crise de gouvernance interne, le comité scientifique est démis et les impérati-
fs commerciaux prennent le dessus. En septembre 2024, OpenAI abolit finalement son
plafond de rentabilité : l'entreprise ne considère plus l'IA comme une science fonda-
mentale au service de l'humanité, mais principalement comme un produit à commer-
cialiser au plus vite.

*Désormais située au cœur d'enjeux économiques, intellectuels, militaires et même philosophiques,
l'intelligence artificielle telle que nous la connaissons aujourd'hui est souvent présentée comme
l'aboutissement sublime et inéluctable du génie humain plutôt que comme la dernière technolo-
gie paradigmatique du capitalisme américain.*

I.2 L'IA des GAFAM aux quatre coins du monde : quels impacts politiques, économiques et sociaux ?

*Une fois ce produit commercialisé et présent - de gré ou de force - dans l'intégralité des objets
numériques de consommation grand public comme des outils professionnels, que se passe-t-il ?*

Brève analyse d'impacts :

I.2.a Sur les individus

« L'assistant IA » : risque de prolétarisation intellectuelle

L'obsession de l'industrie numérique Ouest-américaine pour le développement
de produits individuels, qui ont pour principale caractéristique de chercher à prendre
une apparence humaine, est problématique. Pour employer un système automatique
d'analyse de données, de génération d'image ou de traduction, nul besoin d'avoir une
unique interface « conversationnelle » qui peut également « écrire » des poèmes et paraître
ressentir des émotions, à un point parfois troublant. Leur aspect délibérément trom-
peur, anthropomorphique, ainsi que la façon insidieuse dont ces "assistants IA" sont
rapidement incorporés à tous les produits numériques, ne remplissent aucune nécessité
fonctionnelle, mais correspondent seulement au plan marketing d'entreprises qui font
face à de sérieuses pressions de rentabilité. Souvent utilisées, dans la chaîne de concep-
tion des outils numériques, comme des nouvelles modalités de l'interface homme-ma-
chine, différentes ia risquent de rapidement devenir des intermédiaires numériques

entre individus et sphère publique, il est donc d'autant plus crucial de comprendre et contrôler les effets que cela entraîne sur les individus.

Certains chercheurs évoquent ainsi le risque consistant à déléguer des capacités psychiques ou cognitives et sociales à des dispositifs automatisés, en particulier, avec les IA génératives et les "agents conversationnels" : les capacités d'écriture et de langage risquent d'être affectées, ainsi que les capacités relationnelles¹⁷. Plus généralement, le petit nombre d'acteurs produisant ces systèmes, qui agrègent pourtant des marqueurs culturels forts, laisse craindre un processus de normalisation culturelle, déjà largement à l'œuvre avec les plateformes mondiales de communication et leurs algorithmes de recommandation.

I.2.b Sur la concentration économique

Captation de la valeur ajoutée

Le déploiement unilatéral sans appel de systèmes "utilisant l'IA" à l'ensemble du monde occidental place la focale sur un certain type d'IA, qui répondent à des besoins parfois imaginaires, mais concentrent des capitaux bien réels. Les géants du numérique américains - dont les pratiques monopolistiques ne sont plus à prouver - détiennent un contrôle quasi total sur les technologies concernées mais aussi sur les infrastructures qui la rendent possible. Le développement et l'utilisation de modèles de langage massifs requiert de grandes puissances de calcul et des quantités immenses de données : bien que les besoins de transparence de la communauté scientifique ont jusqu'à présent favorisé l'utilisation de jeux de données ouverts, la puissance de calcul, en revanche, est aux mains d'une poignée d'acteurs industriels dominants. Ainsi, une unique entreprise, NVIDIA, fournit près de l'intégralité des processeurs de dernière génération utilisés pour développer et distribuer les systèmes d'ia.

Captation de la valeur travail

L'IA pousse à son paroxysme la captation de la valeur ajoutée produite par le travail, intrinsèque au fonctionnement du capitalisme, en l'étendant de plusieurs façons.

D'abord en étant fondée sur un nouveau mode de capture de valeur développé par les industries numériques américaine. En s'attribuant l'exploitation sans conditions de l'ensemble des contenus, données, traces numériques produits par les utilisateurs de leurs plateformes, les GAFAM s'approprient une quantité de valeur immense produite en continu par des milliards d'individus. Ce *Digital Labor*¹⁸ fait de nous tous des travailleurs de l'industrie numérique, et la valeur de cette production se cristallise aujourd'hui dans des systèmes d'ia qui sont directement entraînés sur nos données.

¹⁷ *Le capital que je ne suis pas ! Mettre l'économie et le numérique au service de l'avenir*. Anne Alombert, Gaël Giraud, Fayard, 2024

¹⁸ Dimitrova, S. (2016) . Dominique Cardon, Antonio A. Casilli, *Qu'est-ce que le digital labor ?* Bry-sur-Marne, Ina, coll. Études et controverses, 2015, 101 pages. Questions de communication, n° 30(2), 407-409.

Par ailleurs, au quatre coins de la planète - dans les coins plus démunis, c'est entendu - des individus se trouvent exploités par des plateformes de travail digital comme Amazon Mechanical Turk. Ils sont employés pour annoter les données (par exemple, étiqueter des images pour entraîner ensuite des systèmes de génération visuelle), mais aussi pour corriger manuellement les réponses erronées des systèmes d'ia. Au-delà de leurs conditions de travail misérables, des études récentes montrent les conséquences psychologiques désastreuses de l'exposition prolongée à des contenus générés souvent nauséabonds.

I.2.c Sur le salariat

Automatisation des tâches intellectuelles ? ¹⁹

Les premiers systèmes d'ia générative massivement déployés ont été des outils de bureautique : le FMI parle de 40% des emplois actuels qui pourraient être "affectés" à terme (ie. remplacés ou complétés) par l'IA²⁰. Au delà des conséquences socio-économiques d'un tel scénario, se pose la question de la nature de ces emplois, et de leur évolution : seraient remplacées les tâches normalisées à faible valeur ajoutée, laissant la place à d'autres tâches d'autant plus aliénantes qu'elles seraient le lieu d'un rapport inégal entre homme et machine. Des *Bullshit jobs*²¹ aux *AI jobs* ?

I.2.d Sur l'environnement

Intensification des travers du numérique

Parce qu'elle est immensément consommatrice de puissance de calcul, qui se trouve mobilisée dans les data centers de Google, Amazon, Meta ou Microsoft, le développement actuel de l'IA porte un coût environnemental prohibitif. Les estimations de l'Agence Internationale de l'Énergie portent le coût énergétique des data centers à 4% de la consommation mondiale en 2026²², soit l'équivalent du poids énergétique du Japon, pays de 126 millions d'habitants. Cette augmentation de la demande serait en grande partie attribuée à l'IA (l'autre facteur de croissance est la demande en crypto-monnaies). En découlent les conséquences habituelles : émissions de gaz à effet de serre, forts risques hydriques dans les zones autour des datacenters²³, extraction de matières premières et pollutions associées, déchets électroniques.

Il apparaît donc que le développement et le déploiement actuel de l'IA se font selon des modalités porteuses de nombreuses conséquences négatives : sur les individus, sur les équilibres économiques, sur le travail, sur l'environnement.

¹⁹ Pour une réflexion approfondie sur l'évolution du travail à l'ère de l'automatisation des tâches intellectuelles, voir *La Société automatique - 1. L'avenir du travail*, Bernard Stiegler, Fayard, 2015

²⁰ *AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity*. Kristalina Georgieva, Fonds Monétaire International, 2024

²¹ David Graeber, *Bullshit Jobs*, Les Liens qui Libèrent, 2018

²² Source: International Energy Agency, 2024

²³ *The environmental footprint of data centers in the United States*, Md Abu Bakar Siddik et al 2021 *Environ. Res. Lett.* 16 064017

I.3 Un sésame politique

Avec sa promesse de faire germer la conscience dans une machine, ses rêves prométhéens et ses vocables ambigus — à commencer par son nom — l'intelligence artificielle inspire le monde politique.

Petit florilège :

I.3.a Altruisme et accélérationisme efficaces

L'altruisme efficace

Développé aux États-Unis dans les années 2000, l'altruisme efficace entend répondre à la question de l'utilisation optimale des ressources pour le bien commun. Les champions de ce courant de pensée s'estiment tout désignés par leurs capacités intellectuelles, financières et techniques supérieures pour hiérarchiser et résoudre les principaux problèmes humains, au premier rang desquels les risques de pandémie, de guerre nucléaire, et l'apparition d'une « *intelligence artificielle générale* » (IAG), parfois appelée « *singularité* ». Doté d'une définition suffisamment floue pour que les uns le considèrent comme déjà advenu tandis que d'autres l'imaginent émerger d'ici à un demi-siècle, ce système d'IA consciente engendré dans notre océan de données pourrait mener l'humanité à une ère de prospérité universelle, ou la faire disparaître. Ce projet a déjà son école, l'Université de la Singularité à San Francisco, fondée et financée par de grandes figures du mouvement (Peter Diamandis et Ray Kurzweil).

L'accélérationisme efficace :

Plus radical que l'altruisme efficace, l'accélérationisme efficace prône le développement technologique débridé pour parvenir au plus vite à cette entité supra-humaine et faire passer l'espèce à un stade d'évolution supérieur, l'affranchissant ainsi des périls auxquels elle fait face. En attendant, il convient de lever tous les freins réglementaires et éthiques, d'ignorer les questions de propriété intellectuelle ou de respect des données personnelles, et, sans perdre un instant, d'accélérer. Ce technolibéralisme décomplexé justifie la mise sur le marché de systèmes dont on peine à comprendre le fonctionnement et les implications (voir le paragraphe [Sur l'explicabilité, p. 26](#)) — comme ChatGPT, que OpenAI a rendu public sans attendre. On sent poindre là le modèle de société présenté par l'industrie numérique et ses alliés au pouvoir, celui de l'impératif fonctionnel, que le philosophe Marcello Vitali-Rosati décrit comme la « *déclinaison capitaliste de l'impératif rationnel, une rationalité inféodée à la nécessité de produire de la richesse et d'accumuler des marchandises* ²⁴ ». De quoi installer dans l'imaginaire collectif le remplacement de l'humain par la machine comme horizon des transformations socio-économiques et technologiques actuelles.

²⁴ Marcello Vitali-Rosati, *Éloge du bug. Être libre à l'époque du numérique*, Zones, Paris, 2024.

I.3.b La planification écologique

Planification industrielle²⁵

Hier encore rangée sur l'étagère des idées mortes, la planification industrielle opère ces dernières années un retour en force des deux côtés de l'Atlantique. Les élites occidentales la jugent à présent indispensable pour concurrencer le développement asiatique. La planification écologique fait elle aussi son chemin. Des démocrates américains favorables à un « New Deal vert » à la présidente de la Commission européenne, Mme Ursula von der Leyen, chacun entend mobiliser les moyens de la puissance publique et des nouvelles technologies pour organiser la transition vers une économie plus verte — mais toujours libérale.

Cyber-éco-socialisme

En France, la gauche propose d'aligner la production sur les besoins sociaux et les contraintes environnementales²⁶. En son sein, des voix soutiennent la possibilité d'une coordination industrielle fondée sur des systèmes de prise de décision collective, qui tirent parti des technologies informationnelles récentes. Telle que proposée par Cédric Durand et Razmig Keucheyan, « *la politisation de l'économie suppose une dialectique sous contrainte entre l'échelon global et les échelons locaux.*²⁷ », dans laquelle les technologies de l'information auront un rôle central :

Le déploiement de la planification écologique devra s'appuyer sur les technologies de l'information, que ce soit pour prendre la mesure des transformations de la nature, suivre les impacts écologiques des processus de production, modéliser les enchaînements productifs correspondant aux différentes options viables ou encore rompre avec l'isolement des consommateurs dans l'expression de la demande grâce à un usage émancipateur des plateformes²⁸.

On trouve donc ici un ensemble de technologies et d'usages très varié, qui en appelleraient aussi bien à des méthodes d'analyse statistique, d'optimisation industrielle, de partage d'information et de mise en réseau.

I.3.c Numérique et gouvernementalité

Gouvernementalité algorithmique

Les relations des individus à l'État, à la participation politique, à la sphère publique, passent de plus en plus par des outils numériques. Certains chercheurs parlent aujourd'hui de « *gouvernementalité algorithmique* »²⁹ pour désigner la manière dont les injonctions et recommandations algorithmiques influencent et contrôlent les comporte-

²⁵ A ce sujet, voir la note sur l'industrie à venir, publiée par X-Alternative et Intérêt Général

²⁶ « [Notre proposition macroéconomique](#) », programme du Nouveau Front populaire (NFP), juin 2024, p.4

²⁷ Cédric Durand et Razmig Keucheyan, *Comment bifurquer. Les principes de la planification écologique*, Zones, 2024, p.202

²⁸ *ibid*, pp. 245-246

²⁹ Antoinette Rouvroy et Thomas Berns Rouvroy, A. et Berns, T. (2013) . *Gouvernementalité algorithmique et perspectives d'émancipation Le disparate comme condition d'individuation par la relation ?* Réseaux, n° 177(1), 163-196.

ments de manière infraconsciente, constituant ainsi de nouveaux régimes de gouvernabilité et dispositifs de pouvoirs. Apparaissent également de plus en plus nettement des applications directes, verticales et souvent autoritaires d'utiliser des algorithmes décisionnaires : en France, l'algorithme de surveillance des allocataires de la CAF³⁰ ou les nouveaux dispositifs de surveillance déployés pendant les Jeux Olympiques de Paris en sont des exemples.

Démocratie directe

Partant de ce constat, des voies s'élèvent pour utiliser l'IA afin de «faciliter l'expression démocratique directe», souvent sous la forme de plateformes délibératives permettant l'expression des opinions politiques, agencées à des systèmes de traitement de données³¹. Certains imaginent même des «citoyens digitaux»³², des systèmes d'ia répliquant les opinions politiques des citoyens qu'ils représentent, et dotés d'un droit de vote.

I.4 Possibilités d'émancipation et d'organisation collective

Pourquoi, au vue des nombreux impacts négatifs de cette technologie, ne pas simplement appeler à l'arrêt immédiat de l'IA ? Au-delà du fait que les auteurs de cette note se retrouveraient alors probablement sans emploi, il nous semble prématuré d'évacuer la question de la sorte, et peu réaliste à court terme. D'autre part, de nombreuses applications de l'IA demeurent utiles, à condition d'inscrire leur développement dans des processus de prise de décision collective.

Quelles applications ?

L'IA repose sur un ensemble de technologies de traitement de l'information. Dans les grandes lignes, la plupart de ses emplois actuels ont trait à la collecte, la classification, l'analyse et la synthèse d'informations. En particulier, l'IA revêt un réel intérêt industriel dès lors que les volumes de données concernés sont trop importants pour qu'un quelconque traitement manuel soit envisageable : commentaires sur des réseaux sociaux, classification de grandes banques d'images ou de musique, moteurs de recherche, etc.

Ainsi, l'IA apparaît comme une réponse aux surplus d'informations que les technologies numériques libèrent. Dans ce domaine comme dans tous les autres, il nous faut d'abord penser à la sobriété, et évaluer les données qu'il nous est réellement nécessaire de collecter. Une fois ces choix établis, nous pourrons, de la même manière, décider collectivement des systèmes d'ia que nous souhaitons développer. Certains systèmes d'ia nous rendent déjà plusieurs services au quotidien, et il n'est pas exclu qu'elle puisse nous en rendre d'autres dont nous n'avons pas encore idée aujourd'hui. Mais il est essentiel de poser d'abord la question de l'usage que l'on souhaite en faire, il convient de

³⁰ *Profilage et discriminations : enquête sur les dérives de l'algorithme des caisses d'allocations familiales*. Gabriel Geiger, Soizic Pénicaud, Manon Romain et Adrien Sénecat, Le Monde, 04 décembre 2023

³¹ <https://pol.is/home>

³² <https://kof.ethz.ch/en/news-and-events/kof-bulletin/kof-bulletin/2023/12/Supported-democracy-reinventing-direct-democracy-AI-and-voting-twice.html>

replacer au cœur de la conception des systèmes les exigences politiques qu'ils devront supporter.

Prenons pour exemple le projet évoqué précédemment de **Cyber-éco-socialisme**, p. 17 : comme l'affirment à juste titre les auteurs, la planification «ne saurait se réduire à un calcul optimal»³³. Pas question dès lors d'utiliser telle quelle l'IA, qui fait appel en grande partie à des méthodes de Machine Learning, la *science du calcul optimal*.

Quel cadre politique ?

Comme nous l'avons vu, l'IA actuelle est indissociable de la vision politique de ses concepteurs, et de nombreux projets politiques divers se fondent sur son usage, réaliste ou non. L'enjeu pour nous est (1) que ces projets politiques soient débattus, informés et transparents et (2) que le projet politique retenu in fine soit au service de l'intérêt général. Il nous faut pour cela désigner les points d'attention spécifiques qu'un tel projet politique devrait soulever, et donc de décrire comment s'articulent les différentes sphères, scientifiques, industrielles, politiques et sociales qui s'y trouvent mêlées.

Nous allons nous attacher dans la deuxième partie à la description de ce que nous observons actuellement comme étant les principales structures de développement et de contrôle de l'IA. Nous nous apercevons qu'en l'état actuel des choses, le projet politique de l'IA n'est ni débattu et transparent, ni au service de l'intérêt général. Dans la 3e partie, nous formulons des propositions pour changer de voie.

³³ *ibid*, p. 245

II. Points névralgiques : les structures de développement et de contrôle de l'IA

“La culture est déséquilibrée parce qu'elle reconnaît certains objets, comme l'objet esthétique, [...] tandis qu'elle refoule d'autres objets, et en particulier les objets techniques, dans le monde sans structure de ce qui ne possède pas de significations, mais seulement un usage, une fonction utile. Devant ce refus défensif, [...] les hommes qui connaissent les objets techniques et sentent leurs significations cherchent à justifier leur jugement en donnant à l'objet technique le seul statut actuellement valorisé en dehors de celui de l'objet esthétique, celui de l'objet sacré. Alors naît un technicisme intempérant qui n'est qu'une idolâtrie de la machine, et, à travers cette idolâtrie, par le moyen d'une identification, une aspiration technocratique au pouvoir inconditionnel.”

G. Simondon, *Du mode d'existence des objets techniques*, Introduction, p. 10

“Les éducateurs peuvent développer chez l'enfant le respect de la machine en apprenant à l'enfant à la construire, à la réparer, à l'entretenir avant et après son utilisation. [...] Sans doute, il ne faut pas tomber dans une idolâtrie de la machine. Mais entre l'idolâtrie et le mépris existe la saine connaissance fondée sur une fréquentation attentive.”

G. Simondon, « Prolégomènes à une refonte de l'enseignement », in *Sur la technique*, pp. 252-253

Ces quelques mots auront exprimé, nous l'espérons, la nécessité de s'attarder sur des concepts clés de fonctionnement communs à la plupart des systèmes d'IA couramment utilisés

II.1 Sur le fonctionnement d'une ia

De sa conception à sa réalisation, l'intelligence artificielle n'est pas neutre politiquement. Pour éclairer les intrications entre technique et politique qui se logent au sein de la construction d'une ia, il faut ouvrir la boîte noire : comprendre ce dont il s'agit et comment fonctionnent ses mécanismes d'apprentissages.

II.1.a Termes importants et critiques associées

Une ia : courte définition

Une ia est un système entrée / sortie, une fonction mathématique qui apprend, à partir d'une base de données, à fournir les réponses correctes à des questions précises, dans le but de remplir un objectif défini. S'il s'agit par exemple d'identifier des animaux, elle doit prédire correctement l'étiquette associée à chaque image ("chien" ou "pélican"). Pour cela les ingénieurs entraînent le programme à partir d'une banque d'images étiquetées, avec pour objectif de minimiser l'erreur de prédiction. Cette erreur de prédiction est quantifiée à travers une fonction de perte (ou fonction objectif, voir **Exergue 2 : fonctions objectif**, p. 23). Un protocole réajuste les paramètres après chaque tentative et, lorsque la probabilité d'échec devient suffisamment faible, le système peut être distribué.

Fonction entrée/sortie, paramètres, entraînement, données, objectifs

Une ia est donc une **fonction entrée/sortie** qui répond à une tâche précise, sur la base de **paramètres** définis lors d'un **entraînement** sur des **données**. Cet entraînement vise à optimiser une **fonction de coût** (ou fonction objectif) définie selon les **objectifs** que le système vise à remplir. Entrée/sortie, paramètres, entraînement, données, fonction de coût, objectifs : tout cela est défini par l'ingénieur qui conçoit le système. En effet les systèmes d'ia sont des assemblages de ces éléments, et le travail des ingénieurs est en grande partie d'étudier quelles associations fonctionnent pour remplir une tâche donnée.

Ils peuvent jouer sur la structure de la fonction entrée/sortie : quel format de données (image, texte, son) sera absorbé par quel type d'architecture interne (réseaux de neurones, couches denses ou faiblement connectées, activations sigmoïdes, mécanisme d'attention). Ainsi en 2017, l'équipe de Google Brain fait une découverte stupéfiante en cherchant à développer un système de modélisation statistique du langage textuel, à partir du mécanisme dit *d'attention*. Ce mécanisme repose sur une idée simple mais puissante : chaque mot dans une phrase est conditionné par la présence de tous les

autres mots, il est essentiel de prendre en compte l'ensemble du contexte pour chaque mot, c'est ce qu'on appelle *l'attention*. Cette *attention* permet de pondérer les informations et de capturer les relations complexes entre les mots, enrichissant ainsi la compréhension du texte. Ce qui constitue une réelle surprise pour les chercheurs, c'est de constater qu'il n'y a pas besoin d'autre chose : en appliquant ce mécanisme à toutes les parties du modèle, et sans inclure de règles lexicographiques ou d'autres modélisations, le système parvient à un modèle du langage étonnamment précis. Ils présenteront ainsi dans leur article « *Attention is all you need* » l'architecture du transformer, brique fondamentale de tous les modèles d'apprentissage profond modernes. Impossible de dire alors pourquoi exactement ce mécanisme rend possible ce fonctionnement, impossible de le dire encore aujourd'hui.

Pareillement, les ingénieurs choisissent les données d'entraînement qui servent à calibrer le modèle, une fois l'architecture choisie. Systèmes complexes demandant données massives, ils se servent de bases de données bien trop vastes pour être jamais évaluées par un opérateur humain, ni même parfois par des méthodes de représentation statistiques (imaginez chercher à représenter l'ensemble des contenus textuels partagés sur internet). Impossible de prévoir ce qui va sortir de ces machines, mais l'expérience nous a largement prouvé ce que nous pouvions anticiper : ces systèmes de moyennage statistiques reproduisent les biais que portent leurs données. Si les images de docteurs représentent majoritairement des hommes blancs, le système calculera que les docteurs sont des hommes blancs.

Une fois les systèmes entraînés, il serait illusoire de penser que l'étude de leurs paramètres (tels qu'ils sont déterminés par l'entraînement, et donc correspondant à un système qui parvient à remplir sa tâche) pourrait nous apprendre quelque chose. L'étude des structures internes de systèmes d'ia révèlent que leurs représentations internes sont très difficiles à interpréter.

Enfin, la question de la définition arbitraire, et trop peu discutée, des fonctions objectif, nous semble primordiale, et nous l'abordons plus spécifiquement dans [l'Exergue 2 : fonctions objectif, p. 23](#).

II.1.b Domaines de recherche et d'applications

Types d'ia

La vaste majorité des systèmes d'ia aujourd'hui utilisés sont ce que l'on appelle des systèmes de *Machine Learning*, des machines qui peuvent apprendre une tâche à partir de l'exposition à des données d'entraînement. Au sein du Machine Learning, le mécanisme du *Deep Learning* est ce qui permet au système d'apprendre de nombreuses tâches très complexes³⁴.

³⁴ Nous renvoyons le lecteur curieux vers [l'excellent travail de vulgarisation de 3Blue1Brown sur Youtube](#).

Les tâches qu'ils parviennent à accomplir sont nombreuses, on en citera quelques unes, réparties en plusieurs champs :

- *Vision par ordinateur* : classification d'images, de son, génération d'images et de vidéos, synthèse vocale, reconnaissance faciale
- *Traitement du langage naturel* : traduction, analyse de texte, classification, génération de texte (modèles de langage massifs), moteurs de recherche
- *Audio* : transcription, reconnaissance de voix, synthèse vocale, génération de son
- *Statistique* : régressions, classification, prédiction de séries temporelles, ciblage et recommandation de contenus

Domaines d'application courants

Nous listons ici quelques domaines où l'IA est aujourd'hui largement déployée : ciblage publicitaire (GoogleAds), systèmes de recommandation (Meta, TikTok, YouTube), bureautique avancée (ChatGPT, Office360), robotique (Boston Dynamics), armement, renseignement militaire et civil (CAF, DGFIP), finance (Bloomberg, Kayrros, Blackrock).

Exergue 2 : fonctions objectif

Pour pouvoir apprendre une tâche à un système d'ia, il est nécessaire de la formaliser mathématiquement. Cela suppose de pouvoir traduire un objectif qui est parfois flou (eg. présenter des contenus intéressants à des utilisateurs) en une valeur objectivable et, surtout, calculable (eg. le temps passé sur l'application, le nombre de clics).

Les ingénieurs ont souvent bien conscience qu'ils introduisent des biais forts en décidant d'aligner leur système sur une fonction objectif, d'autant plus que ces biais apparaissent parfois de manière spectaculaire, par exemple lorsque des systèmes de recommandation se mettent à favoriser les contenus haineux, ou quand des systèmes de retouche photo transforment Obama en homme blanc (sur ce point, voir [Pourquoi l'IA voit Barack Obama Blanc](#), le Monde Diplomatique, novembre 2024, article publié par des co-auteurs de la présente note).

L'idée est peut-être moins répandue que *le fait même d'adosser un système d'optimisation à une règle de récompense unique entraîne nécessairement des effets normalisateurs et discriminants*. Le paradigme du Machine Learning est fondamentalement une *science de l'optimisation*. Or, pas d'optimisation sans objectif.

Il existe des méthodes d'apprentissage supervisé par retour humain (Reinforcement Learning with Human Feedback) qui visent à adapter une fonction objectif à partir de multiples retours d'utilisateurs, sous forme de notations ou de préférences. Ce principe a, par exemple, été utilisé pour ajuster ChatGPT et réduire les risques que le système ne reproduise des propos problématiques (à caractère misogynes ou racistes, par exemple) provenant de ses données d'entraînement.

Ces méthodes ont l'avantage de ne pas imposer une fonction déterminée à l'avance par les développeurs, permettant ainsi de prendre en compte une variété d'objectifs partiels. Cependant, elles présentent aussi un inconvénient : l'objectif final devient implicite et n'est plus formulé de manière mathématique et transparente, ce qui complique sa compréhension. Il est important de rappeler qu'il s'agit toujours d'une formalisation d'objectif, qui représente un choix parmi d'autres et qui devrait rester ouvert au débat.

Ainsi devons nous insister : *il faut reposer sérieusement la question du choix des fonctions objectif.*

Dans le cas d'un algorithme de recommandation, les choix possibles d'objectifs sont nombreux : qualité du temps passé, connaissances acquises, qualité du débat en commentaires. Mais la manière de les traduire en fonctions objectif n'est pas toujours évidente, et demande d'y consacrer des travaux de recherche³⁵.

Domaines de recherche à investir

Le FAT/ML (Fairness, Accountability, and Transparency in Machine Learning) vise à étudier et atténuer les impacts discriminatoires de l'IA. Les chercheurs.ses ne prônent pas l'interdiction de certaines fonctions objectif spécifiques, mais militent pour une conception d'algorithmes intégrant des considérations d'équité et de transparence. Sont fréquemment critiquées les fonctions objectif traditionnelles, qui se focalisent uniquement sur la précision ou la maximisation des profits, en négligeant les contraintes éthiques.

En somme, les chercheurs de ces domaines appellent à intégrer au fondement même du développement de l'IA les principes de **Fairness, transparence, responsabilité, alignement, protection des données personnelles et sensibles, cybersécurité, et gouvernance.**

Ainsi, la conception d'une ia, processus technique par excellence, est empreinte de nombreux choix effectués par des acteurs qui ont leur propres intérêts (GAFAM) et biais politiques et sociaux. Si ces influences sont inévitables, des gardes fous pourraient être mis en place pour garantir un degré minimum d'alignement avec des objectifs collectifs et transparents décidés en amont. Nous formulons des propositions en ce sens dans la 3e partie.

³⁵ voir par exemple les travaux de l'association Tournesol, qui étudie justement l'usage d'autres fonctions objectif pour les algorithmes de recommandation : <https://tournesol.app/>

II.2 La recherche scientifique

Nous en venons naturellement à poser la question primordiale : comment opère la recherche fondamentale en Intelligence Artificielle ? Nous poserons la question depuis deux points de vue : celui de la qualité de la recherche d'une part, et des choix de carrière qui s'offrent à de jeunes chercheurs d'autre part.

II.2.a La qualité de la recherche et ses évolutions récentes

Vérification des preuves et qualité des articles

Notons tout d'abord une tendance dans le monde académique dans le champ de l'IA à négliger la relecture minutieuse des preuves mathématiques, de l'aveux même des relecteurs, et conformément aux instructions des conférences³⁶. La transparence des relectures dans certaines conférences permet d'en rendre compte, et le nombre de papiers assignés comme les délais de traitement rendent presque impossible une relecture rigoureuse. Il devient courant de travailler avec des systèmes dont les mécanismes internes ne sont pas explicables, l'exemple des transformers (le mécanisme aux fondement des modèles de langage comme ChatGPT) est le plus parlant.

La revue scientifique est naturellement biaisée et en partie arbitraire³⁷: en effet elle juge si la contribution scientifique est "significative", ce qui est une notion subjective. Certains relecteurs pourraient ainsi considérer qu'une avancée dans la performance des modèles de langage est plus "significative" que des travaux permettant d'accroître la sécurité du traitement de données sensibles. Dès lors, il devient plus dur de publier des résultats négatifs en IA, alors que la démarche scientifique se fonde autant sur l'établissement de résultats négatifs que positifs.

Impact des prépublications sur ArXiv

Le recours croissant aux prépublications sur des plateformes comme ArXiv influence fortement la recherche, pas seulement en IA. ArXiv est une archive ouverte de pré-publication (ie. une version d'un article scientifique qui précède son acceptation par le comité de rédaction d'une revue) d'articles scientifiques. ArXiv est très largement utilisée par les chercheurs en IA et constitue la source de référence pour les professionnels cherchant à connaître l'état de l'art du domaine.

Cette pratique encourage la publication rapide pour sécuriser une idée, permettant à un.e chercheur.se de s'assurer la parenté d'un concept si il.elle craint qu'une autre équipe travaille sur le même volet, parfois au détriment de la vérification approfondie et de la qualité. Nous le concéderons cependant, elle permet aussi un accès immédiat aux résultats de recherche, y compris pour des travaux qui ne seraient pas publiés autrement. Les avantages et inconvénients de cette approche sont sujets à débat.

³⁶ "Looking at supplementary material is at the discretion of the reviewers." <https://neurips.cc/Conferences/2024/CallForPapers>

³⁷ voir les travaux menés par la conférence NeurIPS à ce sujet : *The NeurIPS 2021 Consistency Experiment*, Alina Beygelzimer, Yann Dauphin, Percy Liang, and Jennifer Wortman Vaughan, NeurIPS 2021 Program Chairs, 8/12/2021

Néanmoins il est indéniable qu'une certaine priorité est donnée à la nouveauté : la recherche tend à privilégier les annonces spectaculaires et innovantes. Phénomènes de mode, surcharge de contributions peu significatives, saturation de la revue par les pairs, conflit d'intérêt avec les groupes privés, tout cela a pour conséquences le manque de connaissances sur les risques de l'IA, la légitimation de l'usage inconditionnel de l'IA uniquement sur la base de sa spectacularité, l'utilisation déraisonnée de boîtes noires dans d'autres domaines scientifiques.

Enfin, la fréquence de publications et les capacités financières et publicitaires de certains grands laboratoires (Google Deepmind, Meta FAIR) les incitent, dans certains cas, à se passer des revues scientifiques pour publier uniquement sur ArXiv et diffuser leurs travaux avec leurs propres moyens de communication.

Fondements mathématiques

Force est de constater qu'il n'existe peu ou pas de recherches théoriques approfondies sur des innovations récentes en IA, comme les transformers (voir [II.1.a Termes importants et critiques associées, p. 21](#)). De nombreuses publications visent à améliorer leur efficacité de façon marginale, mais relativement peu présentent des recherches théoriques approfondies sur leur fonctionnement. Les réseaux dits *Generative Adversarial Networks*, antérieurs aux transformers et qui ont constitué une avancée majeure du début des années 2010 ont bénéficié de plus de travaux sur leurs fondements mathématiques, qui ont abouti à de premières conclusions plusieurs années après le déploiement massif de systèmes utilisant ce concept. Ainsi la question de l'explicabilité ou du moins de l'interprétabilité des fonctionnements internes de systèmes d'ia n'est pas insoluble, mais il conviendrait peut-être d'établir des règles de prudence et d'étendre les délais (extrêmement courts aujourd'hui) de mise en application de théories fondamentales, en plus d'encourager plus généralement la recherche explicative dans le domaine.

Sur l'explicabilité

Nous l'avons évoqué précédemment, la structure même des systèmes d'ia diminue considérablement, voire invalide totalement, nos capacités à expliquer son fonctionnement interne. Comme le décrit le chercheur Christophe Denis, « Expliquer » le fonctionnement de modèles d'IA, c'est être capable de traduire la réponse du système en une « *série d'étapes reliées entre elles par ce qu'un être humain peut interpréter sensément comme des causes ou des raisons* »³⁸. Ne pas pouvoir expliquer comment fonctionnent ces systèmes pose de sérieuses questions épistémologiques, mais également terriblement concrètes, sur le contenu proprement scientifique de ces technologies.

Et les conséquences vont en cascade : de la recherche pharmaceutique à la biologie génétique en passant par la sociologie ou l'économie, des pans entiers des sciences modernes s'attellent d'ores-et-déjà à employer des systèmes d'ia pour faire avancer les connaissances dans le domaine. Nous posons la question : faisons nous réellement

³⁸ Christophe Denis, « *Esquisses philosophiques autour de la compréhension de phénomènes complexes avec des outils de prédiction basés sur de l'apprentissage machine* », Conférence francophone sur l'extraction et la gestion des connaissances - Atelier Explain'AI, Blois, janvier 2022.

avancer les connaissances, en utilisant des systèmes que nous ne connaissons pas ? Il convient bien évidemment d'éviter les sentences irrévocables, et si le maniement intelligent et justifié de certaines technologies peut faire progresser le traitement de maladies ou l'analyse de documents, nous ne saurions nous y opposer. Mais gardons nous néanmoins de manifester un engouement sans nuances pour ces dynamiques, et encore moins de les encourager³⁹.

II.2.b Les choix de carrières possibles pour un.e jeune chercheur.se

Nous posons la question, devant le constat sans équivoque que les laboratoires privés des GA-FAM parviennent aujourd'hui à capter une grande part des personnes les plus qualifiées en IA, sur les choix de carrière qui s'offrent à un.e jeune chercheur.se.

Attractivité de la recherche publique

Du point de vue des infrastructures matérielles et des ressources mises à disposition, les laboratoires publics sont aujourd'hui bien moins attractifs que leurs équivalents privés. En revanche, le prestige et le rayonnement international de certains laboratoires publics parviennent largement à concurrencer, voire dépasser, dans certains domaines plus théoriques, ceux d'organismes privés. En particulier, les laboratoires de l'INRIA en France ont toujours été et restent des références de niveau mondial dans l'IA.

Les salaires proposés, en revanche, sont bien souvent beaucoup plus élevés dans les laboratoires privés. Le programme Choose France CNRS⁴⁰ propose des postes très bien payés et avec financements à des chercheur.ses en début de carrière, dans le but d'éviter la fuite des cerveaux vers des laboratoires privés ou étrangers. Ce type de programme crée néanmoins des tensions au sein du monde académique, en dérogeant à la règle d'égale rémunération pour tous les domaines, en vigueur au CNRS.

Chaires conjointes industrie / académie

Beaucoup de chercheurs travaillent à la fois pour une institution académique et une entreprise technologique. Un exemple parlant concerne les professeurs du master Mathématiques, Vision, Apprentissage de l'université Paris-Saclay, considéré comme à la pointe de l'enseignement mondial dans le domaine : nombre d'entre eux sont affiliés à des entreprises américaines. Cette double affiliation souligne la porosité croissante entre le monde universitaire et l'industrie.

³⁹ Voir à ce sujet l'édifiante interview de Joëlle Barral, directrice scientifique de Google Deepmind : « *L'IA va permettre d'accélérer la recherche scientifique, bien plus qu'on ne peut l'imaginer* », Le Monde, 12 octobre 2024

⁴⁰ <https://www.ins2i.cnrs.fr/en/cnrsinfo/choose-france-cnrs-ai-rising-talents>

Programmes doctoraux des GAFAM

En école ou dans les programmes doctoraux, l'implantation des laboratoires au plus proche des centres universitaires (Paris, avec l'implantation récente de nombreux laboratoires des GAFAM est un exemple frappant) pose question : la structure des masters exige des stages de recherche souvent réalisés dans des laboratoires affiliés à des industriels, et les doctorats se poursuivent fréquemment selon la modalité des thèses CIFRE, où le financement est assuré par une entreprise privée. Cela n'a pas lieu qu'en Europe : aux Etats-Unis et en Asie, les stages chez les entreprises de la Silicon Valley deviennent un facteur important d'une bonne réussite de thèse. La prise en stage est pratiquement plus valorisée que les publications, le stage devient une sorte de "seconde validation par les pairs". Les pairs en question étant les pontes de l'industrie technologique américaine.

Si l'on pourrait croire au mirage de la liberté parfois accordée en surface aux chercheurs dans ces laboratoires, les risques professionnels et réputationnels à porter une analyse critique sont croissants. Les signaux envoyés par les GAFAM sur ce point sont inquiétants : on pensera à Google, Twitter et Meta qui ont tous licencié leurs équipes de recherche sur l'IA éthique⁴¹. Chez Google, nous citerons l'exemple du chercheur El Mahdi El Mhamdi, à qui l'on a refusé de publier ses travaux sur l'impossibilité de sécuriser les modèles d'IA trop massifs⁴².

II.2.c Comparaison avec d'autres domaines

Aérospatiale, pharmaceutique, tabac

En suivant un parallèle avec les industries pharmaceutique dont un fonctionnement très particulier les rend perméables aux intérêts privés par la mainmise de leurs laboratoires sur la recherche, nous pourrions comparer l'activité des acteurs privés de la Big Tech américaine à celle du lobby du Big Tobacco dans les années 60-70 aux Etats-Unis⁴³. Nous voyons également une comparaison possible avec le phénomène décrit par Irénée Régnauld et Arnaud Saint-Martin, qui voient avec l'arrivée de SpaceX et la mentalité startup une transformation en profondeur de la recherche en aérospatial⁴⁴.

Ces industries, après avoir suivi des trajectoires d'innovation et de développement peu régulés qui ont conduit à des dérives, au moment de leur réappropriation par des grands groupes industriels, ont vu naturellement la puissance publique leur opposer des régulations ambitieuses. Notamment, leurs contraintes à l'innovation repose aujourd'hui sur le principe de non-conformité : par défaut, un produit pharmaceutique ou un aéronef est interdit à la commercialisation, avant que ne soient établies des preuves que son utilisation ne présente pas de dangers, selon des critères pré-établis.

⁴¹ Twitter: [toute l'équipe lors de l'arrivée d'Elon Musk](#). Google en 2020: [2 chercheuses qui co-dirigeaient une équipe de recherche en IA éthique](#). Meta 2023: [dissolution de son équipe d'IA Responsable](#)

⁴² Voir son témoignage : « [Les IA génératives font diversion à celles des réseaux sociaux](#) » sur Usbek & Rica, Pablo Maillé, 9/05/2023, et [son article finalement publié indépendamment sur ArXiv](#).

⁴³ The Grey Hoodie Project: Big Tobacco, Big Tech, and the threat on academic integrity. Mohamed Abdalla, Moustafa Abdalla, 2021

⁴⁴ *Une histoire de la conquête spatiale. Des fusées nazies aux astrocapitalistes du New Space*. Irénée Régnauld, Arnaud Saint-Martin, éditions La Fabrique, 2024

L'introduction de principes de non-conformité pour l'IA correspondrait à une forme de régularisation de l'industrie numérique. Cela permettrait par ailleurs de protéger les consommateurs comme les industriels (non numériques) contre des produits défectueux voire malveillants.

Ainsi, la recherche scientifique en intelligence artificielle est contrôlée et menée par une poignée d'acteurs privés, et semble court-circuiter plusieurs processus élémentaires de la recherche scientifique, participant ainsi à l'accélération tête baissée que promeuvent ces entreprises. Nous formulons des propositions pour renouer avec des pratiques de recherche robustes et éthiques dans la 3e partie.

II.3 Considérations géopolitiques

II.3.a Le poids des acteurs américains :

Modèles, processeurs, cloud

Les modèles phares de l'IA ont majoritairement été développés par des entreprises américaines : OpenAI, Google, Meta, Amazon.

L'industrie des processeurs graphiques (GPU), domaine d'innovation de pointe indispensable au développement et au déploiement des modèles est trustée par une entreprise, NVIDIA qui possède environ 90% des parts de marché. Nvidia est une entreprise américaine, ses usines sont à Taïwan, mais sa chaîne de valeur est très éclatée géographiquement.

Le dernier élément infrastructurel, le cloud (des data-centers décentralisés et accessibles en ligne, qui permettent d'entraîner puis d'utiliser les modèles d'IA sans avoir besoin d'acheter directement les processeurs) est essentiellement géré par Amazon Web Services, Microsoft Azure et Google Cloud Platform.

II.3.b La place de la Chine

Guerre froide 2.0

Le scénario de conflit entre Chine et USA sur l'IA est largement exagéré⁴⁵ - à dessein - mais la Chine pourrait être le seul acteur capable de rattraper l'avance prise par les Etats-Unis. Ses atouts sont la disponibilité des données, l'acceptation d'une population déjà habituée à l'omniprésence de l'IA car habituée à l'omniprésence du contrôle étatique, et la qualité de ses centres de formation scientifiques⁴⁶. Restent les infrastructures, et avec une hausse de 300% des investissements en machines spécialisées, la Chine entend accroître significativement ses capacités de production de processeurs haut de gamme, pour contourner les restrictions américaines⁴⁷.

⁴⁵ Voir Evgeny Morozov, *Une guerre froide 2.0*, Le Monde Diplomatique, mai 2023

⁴⁶ La Chine compte aujourd'hui beaucoup de laboratoires mondialement connus : IDEA, Microsoft Research Asia, ByteDance, Tencent. L'Université Tsinghua est une référence mondiale dans le domaine.

⁴⁷ Voir les chiffres rassemblés par J.Politano dans [The US-China Chip War is Escalating](#), in Apricitas Economics, 17/02/2024

II.3.c. L'UE essaie d'exister :

Initiatives européennes

Certains acteurs émergent dans l'UE, mais sont encore très dépendants de l'industrie américaine : la start-up MistralAI, souvent citée en exemple d'entreprise d'IA française, est en partie détenue par des capitaux américains. L'UE essaie de développer ses propres infrastructures cloud, comme le Scaleway de Xavier Niel ou le cluster Jean Zay pour la recherche académique européenne, dans le but de réduire la dépendance aux acteurs américains.

En France

Beaucoup des talents en IA reconnus au niveau mondial sont français. Un exemple historique est le logiciel Scikit-Learn, développé par l'Inria, simplifiant la programmation de modèles d'ia et utilisé par la majorité des acteurs du domaine dans le monde entier. Plusieurs personnalités françaises importantes dans le domaine travaillent dans des entreprises américaines : Yann Le Cun (inventeur de la backpropagation au cœur du deep learning), ainsi que de nombreux ingénieurs de Meta derrière LLaMA (un des plus performants modèles de langue open-source, un système similaire à ChatGPT).

Ces dernières années, suivant l'impulsion de Meta en 2015, plusieurs entreprises de la Big Tech américaine (Meta, Google, Apple, Amazon) mais également chinoise (Huawei) ont installé des laboratoires de recherche en IA à Paris. L'État finance en partie ces laboratoires, via des chaires, le Crédit Impôt-Recherche, ou le programme France 2030. Ces laboratoires sont avant tout une manière pour ces entreprises de capter les chercheurs fraîchement diplômés des grandes universités européennes en finançant des thèses qui, si elles sont bien mieux rémunérées que dans le public, restent très en deçà des salaires américains. D'une certaine façon, on peut parler de délocalisation pour ces entreprises qui s'installent en France, dans un domaine qui reste extrêmement coûteux en techniciens hautement qualifiés.

II.4 Législation et contrôle de l'État

Nous passerons ici brièvement en revue les règlements existants permettant d'encadrer le déploiement de systèmes d'ia sur le marché européen. Leur critique n'est pas l'objet de la présente note, bien qu'il faille probablement reconsidérer le cadre législatif dans lequel pourrait être mise en oeuvre une politique industrielle ambitieuse (voir partie III.5 Poursuivre le mouvement de régulation pour favoriser une approche européenne du domaine, p. 38).

Règlement sur la Protection des Données Personnelles, 2016

Le RGPD, voté en France en 2016, donne un cadre à l'utilisation des données personnelles mais est dans les faits relativement peu efficace pour encadrer l'IA : l'article 22 interdit la prise de décision sans contrôle humain lorsqu'elle "affecte significativement" la personne, ce qui est peu utilisé en pratique et comporte de nombreuses dérogations. Des contraintes existent pour tout traitement de données personnelles (dont les systèmes d'IA) mais sont souvent non respectées en pratique et non contrôlées. Le RGPD pose le principe d' "accountability" des concepteurs d'algorithmes mais en IA, le grand nombre de données traitées dilue les responsabilités.

Digital Services Act, 2022

Au niveau européen, le DSA est une tentative de régulation des réseaux sociaux et de leurs algorithmes : lutte contre les contenus illicites ; obligation de transparence sur les algorithmes de fléchage des publicités et de proposer un système de recommandation de contenus non-fondé sur le profilage, interdiction de la publicité ciblée pour les mineurs ; responsabilisation des acteurs sur les contenus qui sont mis en ligne et les risques qu'ils génèrent.

AI Act, 2024

L'AI Act, formellement adopté par l'UE début 2024 et en cours de promulgation, est une des réglementations les plus ambitieuses sur l'IA et introduit : l'interdiction de certains systèmes jugés trop nuisibles (exemple : identification biométrique à distance dans les lieux publics en temps réel - même si des dérogations existent) ; des obligations pour les IA à haut risque (systèmes utilisés dans le système éducatif, la gestion RH...) de fournir à un organisme certificateur des informations sur la manière dont l'IA a été entraînée et sur la représentativité des données.

Ainsi, malgré de premières tentatives timides, la législation sur l'IA obéit dans son ensemble aux préceptes de l'accélérationnisme : limiter la régulation pour garantir le "progrès". Mais alors, qui définit la direction du progrès ? Sans couper court à toute innovation scientifique, le régulateur ne peut-il pas l'orienter dans une direction choisie et délibérée ? Encore faut-il que la multiplicité des ia possibles lui soit donnée à voir. Nous formulons des propositions en ce sens dans la 3e partie.

III. Propositions : pour une politique industrielle des automates computationnels

Forts de ces éclairages, nous pouvons nous atteler à la tâche que nous nous sommes donnée : proposer les fondements d'une nouvelle politique scientifique, industrielle et sociale de l'Intelligence Artificielle. La description des conditions de production des ia, de la connaissance scientifique en IA et des premières initiatives de législation montrent la mainmise absolue d'une vision accélérationniste et capitaliste : notre proposition politique devrait donc avoir pour but fondamental de permettre l'émergence d'alternatives. Dans la continuité des propositions faites dans un travail précédent sur le numérique, la souveraineté technologique nous paraît être le point d'achoppement de toute proposition de renouvellement démocratique de notre espace numérique.

Si, bien sûr, l'IA n'est qu'un fer de lance, voire un mirage prométhéen, et les efforts à fournir se situent avant tout au niveau de notre stratégie industrielle et des conditions de notre souveraineté numérique, nous nous permettrons ici de nous focaliser sur l'Intelligence Artificielle. D'une part parce que, sur les points ayant trait aux politiques industrielles et au numérique, nous pouvons renvoyer le lecteur vers nos travaux existants sur ces sujets⁴⁸.

D'autre part parce que nous considérons qu'il est sain, en des temps où nos imaginaires peinent à nous inspirer créativement et politiquement, de s'attaquer à la flèche tout en haut de l'édifice : imaginer de nouvelles bases politiques pour le projet techno-capitaliste quintessentiel, c'est affirmer sans ciller et au risque de déplaire à certains, qu'il y aura, toujours, des alternatives.

Formulons donc quelques propositions. Leur vocation est avant tout d'ouvrir le débat, et non de le conclure. Notre principale préoccupation porte, en plus de la direction actuelle de l'IA et de ses nombreux impacts négatifs, sur l'absence de la possibilité même d'alternatives à l'IA telle qu'elle se déploie aujourd'hui. Nos propositions ont donc comme fonction première de montrer la possibilité de faire autrement. Il va de soit qu'il est possible de faire encore autrement (voir par exemple les propositions d'Usbek et Rica en juillet 2024⁴⁹). L'important est de pouvoir en discuter.

Nos propositions visent donc à répondre à deux problématiques : comment gérer et rationaliser le déploiement industriel de l'IA dans et par la sphère publique ? Comment garantir la multiplicité de ses mises en application dans une inscription sociale, démocratique et émancipatrice ?

⁴⁸ voir notre Note sur la souveraineté numérique, et celle sur l'industrie publiée conjointement avec Intérêt Général : Construire une planification industrielle à partir des besoins.

⁴⁹ «Intelligence Artificielle, 10 propositions pour reprendre le contrôle», Usbek & Rica n°44, été 2024

III.1 Définir collectivement le projet politique qui doit orienter le développement de l'IA

Redonner sens et prises au débat public

Le point de départ doit être l'éducation de tous au numérique, à la fois pour rendre plus accessibles ses usages, mais aussi - et c'est particulièrement important pour l'IA - pour expliquer la manière dont fonctionnent les technologies numériques. Pour nourrir la « *saine connaissance fondée sur une fréquentation attentive* » chère à Simondon, et parce que la compréhension des systèmes techniques permet la décision et l'action collective. Développer des actions équivalentes à celles menées pour donner conscience des enjeux climatiques (les fresques du climat), comme par exemple les Cafés IA que propose partout sur le territoire le Conseil National du Numérique⁵⁰.

Envisager la puissance publique comme un principe d'association entre les pouvoirs politiques, les associations de citoyens, les puissances économiques privées et tous les membres d'une société en général, associés au service d'un bien public vers lequel convergent les intérêts de tous⁵¹. En somme, faire activement démocratie, entendue comme « *l'action qui sans cesse arrache aux gouvernements oligarchiques le monopole de la vie publique et à la richesse la toute-puissance sur les vies* »⁵².

Les automates au service de la puissance publique et des communs

Il importe d'affirmer à nouveau le rôle que peut jouer l'État pour encourager et soutenir la formation de communs qui sont consubstantiels aux technologies numériques. Un commun numérique peut être défini ainsi :

« *une communauté de producteurs et d'utilisateurs qui gèrent une ressource numérique en vue de son enrichissement dans le temps à travers des règles de gouvernance conjointement élaborées et dont ils protègent le libre accès face aux tentatives d'appropriation exclusive* »⁵³.

La pensée des communs numériques est fondamentale pour appréhender des systèmes d'ia qui sont rendus possibles par ces grandes masses de données que plusieurs décennies d'usage d'internet par des milliards d'êtres humains ont accumulées. L'IA peut s'appuyer sur ces communs autant qu'être un appui à leur constitution : l'entretien d'un grand commun numérique comme Wikipédia pourrait être assisté par des systèmes d'ia⁵⁴.

⁵⁰ voir à ce sujet la présentation des cafés IA sur le site dédié : <http://cafeia.org/>

⁵¹ Voir Bernard Stiegler & Ars Industrialis, Réenchanter le monde. *La valeur esprit contre le populisme industriel*, Flammarion, 2006.

⁵² Jacques Rancière, *La haine de la démocratie*, La Fabrique, 2005, p.105

⁵³ Shulz, Sébastien. « *De l'adoption au rejet d'un commun numérique pour transformer la frontière entre État et citoyens. La trajectoire de la Base Adresse Nationale entre contribution citoyenne, autogouvernement et État-plateforme* », Réseaux, vol. 225, no. 1, 2021

⁵⁴ Petroni, F., Broscheit, S., Piktus, A. et al. *Improving Wikipedia verifiability with AI*. Nat Mach Intell 5, 1142–1148 (2023)

Amélioration plutôt qu'augmentation

Pour que l'IA puisse bénéficier à la puissance publique, la développer non pas pour imiter l'esprit humain et aliéner encore des agents territoriaux⁵⁵, mais pour étendre au contraire les capacités de nos services publics. Inscire le déploiement de l'IA par la commande publique dans une volonté *d'amélioration* plutôt que *d'augmentation*. Nous rejoignons Evgeny Morozov⁵⁶, qui empreinte à l'informaticien utopiste Warren Brodey, pour dénoncer les technologies qui visent l'augmentation, comme une simple automatisation des capacités humaines, plutôt que l'amélioration, comme approfondissement des interactions humaines et de la compréhension de la vie.

Et s'il s'agit de penser une « *démocratie numérique* », dans laquelle des intermédiaires numériques et algorithmiques permettraient de renforcer le débat et d'étendre le politique, il est primordial d'y associer la recherche en informatique et celle en sciences sociales.

Ancrer tout le système technique dans un projet politique

Recherche fondamentale : recherche d'abord en épistémologie et en sciences sociales, pour évaluer les apports des automates, leurs limites, et comment leurs systèmes reconfigurent nos accès aux savoirs et nos rapports au politique. Recherche en sciences mathématiques et physiques pour concevoir les modèles, les processeurs, les terminaux.

Asseoir la recherche dans un cadre de respect des données sensibles, de représentation des biais, de sécurité, d'explicabilité ou de priorité à la gouvernance sur la performance. Insister sur cela au moment des allocations de budgets, des recrutements, des évaluations, des promotions, des communications externes.

Recherche-développement : ouvrir des laboratoires de recherche-développement non affiliés à des grands groupes américains, capables de fournir des débouchés technologiques souverains aux citoyens européens, et suivant des intérêts publics et démocratiques plutôt que des intérêts privés et des objectifs de rentabilité. Les grandes entreprises d'intérêt national, dans lequel l'État a des participations (SNCF, EDF, Thalès), pourraient être des structures adéquates pour accueillir ces labos de R&D.

Infrastructures publiques : financer le développement de centres de stockage de données et de puissance de calcul capables de soutenir matériellement les laboratoires de recherche, les associations et les citoyens. Service public⁵⁷, associations, entreprises stratégiques, industries non-numériques : dans cet ordre, faire bénéficier des développements de l'IA à tous ces acteurs.

⁵⁵ Voir le système Albert, conçu par la DINUM pour être utilisés par les agents de France Service : <https://www.info.gouv.fr/actualite/ia-connaissiez-vous-albert>

Nous ne disposons pas d'assez d'informations pour évaluer ce système, mais tout porte à croire qu'il a été construit suivant le paradigme de "l'assistant IA", que nous critiquons fortement (cf I.2.a).

⁵⁶ voir Evgeny Morozov, *Une autre intelligence artificielle est possible*, Le Monde Diplomatique, Août 2024

⁵⁷ Voir par exemple [l'utilisation de l'IA par l'AP-HP](#), ou [le lancement du Datalab de l'Etat](#)

III.2 Prendre des mesures d'alignement et de restructuration de la recherche publique et privée

III.2.a Orienter les jeunes chercheur.ses

Un trou dans la raquette

Au sein des cursus scientifiques, intégrer l'enseignement des sciences sociales, de l'épistémologie, éduquer aux conséquences sociales et économiques de la recherche en IA, prendre exemple sur le département THS de l'Université de Compiègne⁵⁸. Inclure ces réflexions dans les parcours académiques, et mobiliser des associations pour porter une pluralité de discours auprès des étudiants.

Procéder de même pour les cursus de sciences humaines, pour en finir avec l'interdiction de principe de tout discours sur les sciences dures, et avec certaines idées fausses naissant de la méconnaissance technique.

Offres d'orientation : les forums métiers

Développer des alternatives aux stages, alternances et thèses CIFRE dans les centres de R&D privés. Aujourd'hui, les alternatives sont principalement des stages dans des laboratoires publics, qui peuvent sembler moins opérationnels et donc moins attractifs pour des étudiants. Il devrait être possible de développer des offres de stages et thèses sur des sujets appliqués, au sein d'autres structures publiques que le milieu académique. Tirer ainsi parti de ce que les pouvoirs publics ont d'unique : la possibilité de travailler sur des applications stimulantes (lutte contre la désinformation, démocratie participative, accessibilité, etc).

III.2.b Orienter les financements

Financer la recherche publique

La priorité doit être, en matière de recherche, de préserver les institutions publiques de l'influence des intérêts privés. Celles-ci sont les derniers bastions de l'esprit universitaire, qui doit demeurer celui d'un espace hors des affairments du monde extérieur, où l'esprit se donne l'opportunité d'accomplir des œuvres au temps long. Pour les adeptes de l'efficacité, affirmer sans faiblir qu'il ne peut y avoir de science digne de ce nom sans préserver l'indépendance matérielle, temporelle, culturelle, intellectuelle des institutions académiques. Cette proposition nous paraît *self evident* : il faut financer la recherche publique.

⁵⁸ <https://www.utc.fr/formations/diplome-dingenieur/les-formations-transverses-technologie-societes-humaines/tsh-une-pedagogie-innovante-et-une-recherche-pluridisciplinaire/>

Structures de recherche indépendantes

Mis à part la recherche publique, favoriser également l'émergence de structures de recherche indépendantes ou hybrides, à titre d'exemples :

[Dair Institute](#) : fondé par Timnit Gebru, institut de recherche autour de l'éthique et impact sociétal de l'IA, non affilié à une université ou un acteur industriel.

[Conférence FAccT](#) : conférence dédiée à une approche transdisciplinaire de la Fairness, Accountability and Transparency (bien que les représentants des Big Tech y soient de plus en plus présents)

[AI demos lab](#) : laboratoire de recherche sur les pratiques croisées de l'IA et de la démocratie

Mutualiser les capacités de calcul

Le supercalculateur Jean Zay, du CNRS, est un exemple de mutualisation des ressources au niveau national. Son mode de financement et de partage pourrait inspirer d'autres initiatives nationales voire européennes. Aujourd'hui, la plupart des laboratoires publics se dotent de leurs propres processeurs en utilisant leurs fonds courants, ce qui leur garantit une indépendance mais ne permet pas d'assurer l'efficacité des ressources employées. Nous pourrions garantir une meilleure efficacité et une plus grande autonomie en attribuant des fonds spéciaux et en mutualisant les commandes, en ouvrant à des appels d'offres favorisant des acteurs européens plutôt que des industriels extérieurs.

III.3 Engager l'ensemble de la chaîne industrielle

Pour les mesures visant à aligner l'ensemble de l'appareil industriel sur des décisions politiques prises collectivement, ce qui n'est pas spécifique à l'IA mais devrait imprégner toute politique industrielle, voir [Exergue 3 : Recommandations de notre note sur l'industrie, p. 37](#). Nous insistons ici sur un point, relatif aux types d'applications de l'IA qu'il s'agit de favoriser.

IA pour les entreprises et l'industrie

Favoriser les applications industrielles plutôt que l'IA pour les consommateurs, comme le font déjà plusieurs startups européennes (MistralAI ou Aleph Alpha). L'IA comme instrument ou comme outil, mais pas comme produit ou comme service. On note la différence en se demandant dans quel dispositif est utilisée l'IA, par exemple :

- Les systèmes d'ia pour l'optimisation de la gestion des flux (ressources hydriques, livraison et stockage de marchandises, microgrid intelligentes) sont des dispositifs industriels hautement performants, maintenus par des ingénieurs ayant une fine connaissance de leur fonctionnement. Ce sont des *instruments*.
- Les systèmes d'ia pour la traduction ou la transcription de texte (Deepl, Whisper) sont des dispositifs grand public spécialisés, qui permettent d'effectuer des tâches précises. Ce sont des *outils*.
- Les systèmes d'ia conversationnels sans fonction précise (ChatGPT, Perplexity, Claude) sont des dispositifs grands publics au fonctionnement volontairement opaque et à l'apparence travaillée pour rendre dépendant. Ce sont des *produits*. Ces derniers sont à décourager, dans la mesure où leur utilisation présente un caractère aliénant pour l'utilisateur (voir I.2.a).

Exergue 3 : Recommandations de notre note sur l'industrie

Dans la mesure où les questions industrielles relatives à l'IA ne présentent pas de particularités qui nécessiterait un traitement spécifique, nous renvoyons le lecteur à notre note sur l'Industrie, publiée avec Intérêt Général : *Construire une planification industrielle à partir des besoins*, à paraître en Novembre 2024.

Les principales recommandations tirées de cette note qui nous paraissent s'appliquer particulièrement à l'IA sont les suivantes : une politique industrielle doit

1. permettre à l'industrie d'accéder à un savoir fondamental et des moyens de mesure, de calcul, de caractérisation de pointe lorsqu'elle en a besoin.
2. permettre aux chercheurs de trouver une application de leurs recherches fondamentales dans des problématiques socio-économiques concrètes. Mettre en place un système de licences pour faire bénéficier à l'industrie des développements portés par la recherche publique. Éviter à tout prix la dynamique actuelle, qui rapproche les laboratoires du modèle des incubateurs à startups, favorisant le spectaculaire, le rapide, le rentable, plutôt que la recherche qui fait avancer la science.
3. intégrer à la stratégie industrielle les matières premières, l'industrie des processeurs, du cloud, des réseaux, les start-ups innovantes et enfin les produits de consommation.
4. plafonner les aides publiques, renforcer le fléchage des aides vers les PME/TPE pour ré-équilibrer la répartition des bénéficiaires, mieux contrôler l'utilisation des crédits, conditionner les crédits sur des recherches stratégiques pour la planification écologique.

III.4 Aligner les financements publics

*Nous décrivons ici simplement le principal mécanisme actuel d'investissement public dans l'innovation, pour illustrer les montants et les priorités que celui-ci comprend. Pour France 2030 comme pour n'importe quel plan d'investissement public, l'essentiel est de conditionner les attributions de fond aux mêmes priorités industrielles qui auront été décidées collectivement. Pour une description plus générale du fléchage des aides publiques, voir **Exergue 3 : Recommandations de notre note sur l'industrie**, p. 37.*

France 2030

Le plan d'investissement comprend un volet IA, qu'il s'agit d'orienter vers la recherche et le développement d'automates computationnels au modèles alternatifs. Aligner les mesures existantes sur les points précédents, celles-ci comprennent :

- L'appel à manifestation d'intérêt « IA-Cluster », doté de 500 M€, qui vise à faire émerger 5 à 10 universités et écoles en tant que leaders européens et mondiaux en IA.
- Le dispositif IA Booster France 2030 qui aide les PME et ETI françaises dans leurs processus de transformation numérique recourant à l'IA.
- L'appel à projets de 40 millions d'euros pour constituer et mettre à disposition des « communs numériques » pour l'IA générative (bases de données d'apprentissage, IA pré-entraînées, outils d'évaluation des IA, etc.)

III.5 Poursuivre le mouvement de régulation pour favoriser une approche européenne du domaine

RGPD, DSA, DMA

Sur les critiques des régulations existantes et les possibilités de les étendre, nous renvoyons à des travaux existants : une nouvelle version RGPD pour notamment inclure des capacités d'action collective⁵⁹, les appels à intégrer les coûts matériels, environnementaux et sociaux à la régulation⁶⁰, les critiques formulées à l'égard du DMA⁶¹.

Transparence et audit

Renforcer les obligations de transparence des algorithmes d'IA (comment sont-ils entraînés, avec quel objectif?) via des obligations d'audit des algorithmes, et développer au sein de l'UE une autorité interne d'audit des algorithmes.

Algorithmes de recommandation

Instaurer le droit à modifier ou remplacer les algorithmes de recommandations des plateformes en choisissant leur objectif d'optimisation (par exemple, maximiser l'exploration de morceaux inconnus plutôt que le temps d'écoute sur une plateforme de streaming)⁶².

⁵⁹ <https://www.thedigitalnewdeal.org/rgpd-acte-ii/>

⁶⁰ <https://cnnumerique.fr/le-defi-dune-regulation-de-lintelligence-artificielle>

⁶¹ <https://shs.cairn.info/revue-red-2021-2-page-114?lang=fr#sin4>

⁶² Voir les travaux du Conseil National du Numérique à ce sujet :

<https://cnnumerique.fr/nos-travaux/votre-attention-sil-vous-plait-quels-leviers-face-leconomie-de-lattention>

III.6 Proposer une alternative aux discours médiatiques actuels

Regagner du terrain médiatique

Diffuser des narratifs qui proposent d'autres terminologies et d'autres visions du développement technologique, telles que celles portées par X-Alternative, au service de la planification à long terme et de l'intérêt général, ou bien encore celles d'Ars Industrialis, au service de l'intelligence collective et de l'enrichissement des savoirs.

Fournir des éléments de contexte pour contrebalancer les affirmations rarement contredites de personnalités porteuses des messages clés de la tech américaine. Pour cela, il est nécessaire de déconstruire les postures marketing suggérant une "magie" à ces systèmes d'IA, en tâchant d'expliquer leur fonctionnement statistique déterminé et limité.

III.7 Investir et promouvoir une conception démocratique des IA

Participations ouvertes

La volonté d'étendre le contrôle du développement de l'IA à d'autres acteurs, que nous soutenons fortement, suppose de développer les systèmes d'ia en ayant à cœur de permettre leur modification, leur réemploi, leur altération dans des processus ouverts. Cela suppose donc des recherches et des développements techniques dès la conception des modèles.

Cette approche peut grandement bénéficier au développement des modèles, en plus de les ouvrir partiellement à des contributions multiples. L'entreprise OpenAI elle-même appelle les utilisateurs de ses technologies à y contribuer "démocratiquement" dans le but d'améliorer la représentativité des modèles⁶³. Si leurs méthodes restent fondamentalement obscures (set de données privés, modèles boîtes noires) et ne sont donc pas à suivre, nous proposons d'autres approches d'une construction démocratique de l'IA.

Cela peut passer par :

- la constitution de protocoles pour orienter le fonctionnement d'une ia⁶³,
- l'agrégation de choix individuels pour constituer une politique commune⁶⁴,
- la constitution de bases de données textuelles brutes adossées à des systèmes de traitement statistiques⁶⁵,
- la participation communautaire à la politique éditoriale d'un réseau social⁶⁷,
- les plateformes de vote traditionnelles⁶⁸.

Ces approches peuvent être envisagées comme des systèmes de décision à part entière (et donc s'apparenter aux recherches en démocratie numérique), ou bien simplement des méthodes pour orienter consciemment le fonctionnement d'une ia. Dans les deux cas, ils nous paraissent porteurs de possibilités nouvelles pour faire remonter des choix politiques individuels jusque dans le fonctionnement interne de système algorithmiques.

⁶³ <https://openai.com/index/democratic-inputs-to-ai/>

⁶⁴ <https://dl.acm.org/doi/10.1145/3359283>

⁶⁵ <https://arxiv.org/abs/1709.06692>

⁶⁶ <https://pol.is/home>

⁶⁷ voir les *Twitter community notes*, ou le projet participatif Tournesol : [Tournesol.app](https://tournesol.app)

⁶⁸ <https://make.org/FR>

III.8 Normaliser les pratiques industrielles pour minimiser les impacts néfastes et renforcer la cybersécurité.

Principe de non-conformité

L'introduction de principes de non-conformité pour l'IA correspondrait à une forme de régularisation de l'industrie numérique. Cela permettrait par ailleurs de protéger les consommateurs comme les industriels (non numériques) contre des produits défectueux voire malveillants. La forme qu'une telle régulation prendrait, aussi bien techniquement (à quel moment du développement des modèles contrôler la conformité ? quelle différence entre modèles de recherche et modèles commerciaux ?) que politiquement (quelle entité se chargerait de l'audit ? de l'évolution des normes ?), reste bien sûr à débattre.

Conclusion

Nous avons souhaité par cette note proposer un éclairage qui condense nos positions multiples, en tant qu'ingénieurs, chercheuses, philosophes travaillant au sein du large domaine de l'Intelligence Artificielle. Ce travail collaboratif nous a conduit à interroger les fondements épistémologiques, historiques et politiques d'un champ technologique si mal nommé qu'il nourrit des imaginaires qui le dépassent.

Nous sommes ainsi tombés sur de nombreux sujets d'inquiétude, notamment sur les impacts à prévoir de son extension illimitée, hautement profitable aux yeux d'un certain capitalisme n'ayant pas peur de la chute. Prolétarianisation des individus, concentration économique, aliénation du travail, désastre environnemental : le futur de l'IA inquiète. Et le plus inquiétant demeure son uniformité : nul ne semble en mesure de faire émerger d'autres perspectives de développement de l'IA que celles offertes par l'industrie numérique américaine.

Il nous a ainsi semblé nécessaire, parce que ce projet scientifique et technologique parvient à nourrir des projets politiques de tous bords, et parce que malgré tous ses défauts il demeure porteur de certaines possibilités d'émancipation et d'organisation collective, de chercher un échappatoire à ses tendances historiques et actuelles.

Il nous a donc fallu étudier son fonctionnement, identifier ses composantes idéologiques. Considérer la manière dont se déroule la recherche académique, et ses intrications avec des considérations économiques et politiques. Tout cela pour comprendre où agir.

Et nous en sommes venus à nos propositions. Car notre démarche s'inscrit résolument dans la capacité à faire émerger des alternatives. Etant donné la mainmise actuelle d'une poignée d'acteurs sur son développement, qui ne partagent pas ce but, cela suppose de refonder une politique scientifique, industrielle et sociale de l'Intelligence Artificielle. Nous espérons ainsi ouvrir le débat à l'importance d'inscrire, dans l'ensemble de nos choix scientifiques et industriels, la prééminence du politique.

Cela ne concerne pas uniquement l'IA, mais il n'y a pas meilleur terrain que la *science de l'optimisation* pour affirmer avec joie qu'il y aura, toujours, des *alternatives*.
