

ROBOTS VS ALGORITHMES

Prophétie et critique dans la représentation médiatique des controverses de l'IA

[Maxime Crépel](#), [Dominique Cardon](#)

La Découverte | « Réseaux »

2022/2 N° 232-233 | pages 129 à 167

ISSN 0751-7971

ISBN 9782348073809

DOI 10.3917/res.232.0129

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-reseaux-2022-2-page-129.htm>

Distribution électronique Cairn.info pour La Découverte.

© La Découverte. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

ROBOTS VS ALGORITHMES

Prophétie et critique dans la représentation médiatique
des controverses de l'IA

Maxime CRÉPEL
Dominique CARDON

Par vagues successives, au début des années 1960, des années 1990 et au milieu des années 2010, les promesses et les risques de l'intelligence artificielle sont venus occuper de façon spectaculaire la discussion publique (Russell *et al.*, 2010 ; Martin, 1993)¹. Les débats contemporains autour des robots tueurs, de l'automatisation du marché du travail ou de la vie avec des êtres artificiels ont déjà occupé une place importante dans la discussion publique lors des deux premières vagues de promesses de l'IA (Crevier, 1997 ; McCorduck, 2004). Experts, éthiciens, spécialistes du travail, de l'armée ou de la médecine, acteurs politiques et journalistes ont chaque fois initié un large débat public associant des observations contemporaines sur les transformations des technologies informatiques à un ensemble de considérations prospectives mêlant l'économie, le travail et la morale à une réflexion philosophique sur le devenir de l'humanité. Immédiatement soutenue par les investisseurs, les pouvoirs publics, le secteur de la défense, les mondes de l'innovation industrielle et les cabinets de consultants, chaque avancée réalisée dans les laboratoires de recherche cristallise un faisceau d'attentes et de promesses faisant de l'intelligence artificielle une technologie pivot permettant à nos sociétés de s'appuyer sur des machines pour se libérer d'un nombre considérable d'activités (Brynjolfsson, McAfee, 2014 ; Lee, 2018 ; Tegmark, 2017). Cette promesse n'est jamais pensée comme une étape incrémentale dans le développement des systèmes informatiques mais comme un point de bascule brutal vers une nouvelle forme d'organisation de la société (Vinsel et Russell, 2020). À la différence d'autres technologies, les innovations de l'IA agissent comme une sorte de bombe à fragmentation attirant immédiatement l'attention des médias, des régulateurs et des prospectivistes. Une des principales raisons de cette attractivité tient à l'imaginaire

1. Cette recherche a été réalisée avec le soutien de la Chaire « Good In Tech » placée sous l'égide de la Fondation du Risque en partenariat avec l'Institut Mines-Télécom et Sciences Po. Ces travaux ont pu être réalisés grâce au soutien précieux de Jean-Philippe Cointet, Salomé Do et Yanniss Bouachera pour effectuer les traitements de données. Les cartographies et les données présentées dans cet article sont accessibles sur un site permettant de naviguer de façon dynamique dans les visualisations : <https://medialab.github.io/carnet-algopresse/#/publication/fr/>, consulté le 3 mars 2022.

prométhéen de la mise au monde d'une vie artificielle qui l'accompagne sous des habits mythologiques (Mosco, 2005). Les questions éthiques, la demande de régulation des entités artificielles ou l'interrogation sur la possible autonomie d'une nouvelle société d'artefacts se déploient sur un fond d'imaginaire qui a été très profondément alimenté par la littérature et le cinéma de science-fiction. L'IA constitue un cas très particulier de « mythologie » technologique qui a installé l'idée d'un dépassement de l'humanité par elle-même, une libération des contraintes présentes traçant la route vers un futur inéluctable dont la caractéristique est d'être à la fois émancipateur et apocalyptique (Natale et Ballatore, 2020 ; Singler, 2019). L'imaginaire associé aux technologies, surtout lorsque leur déploiement est aussi dense, intense et généralisé que celui de l'IA, n'est pas sans conséquences sur les acteurs qui la fabriquent, la promeuvent ou souhaitent la réguler. Il ne peut être ignoré par les approches des *sciences and technology studies* (Jasanoff, 2015 ; Flichy, 2001).

Cet imaginaire a été à nouveau réveillé, à partir de 2013, avec la troisième vague de l'IA déclenchée par les performances nouvelles des techniques de *deep learning*. Même si ces technologies d'inspiration « connexionniste » ne partagent pas grand-chose de commun avec les approches « symboliques » des deux précédentes vagues (Cardon *et al.*, 2018 ; Metz, 2021), elles ont cependant été identifiées comme relevant du domaine de l'intelligence artificielle. La réintégration des techniques connexionnistes au sein d'un espace de recherche dont elles avaient été précédemment exclues témoigne de la force des résonances symboliques dont le terme intelligence artificielle est investi par les chercheurs de ces différents domaines, mais aussi par l'ensemble des acteurs qui ont intérêt à constamment renouveler la promesse de nouveaux futurs technologiques. Cette nouvelle visibilité de l'IA est inséparable de la réanimation d'un discours critique. Cette exceptionnelle proximité des technologies dites d'intelligence artificielle avec les centres d'intérêt des journalistes, les scénarios des prospectivistes et l'imaginaire du public mérite examen. Dans cet article, nous mettons en évidence la manière dont la nouvelle vague de l'IA mobilise deux représentations critiques contrastées. La première réanime et relance la figure du robot autonome traditionnellement attachée à l'histoire de l'IA. Mais une autre figure critique, les algorithmes, a aussi émergé pour caractériser de nouveaux risques beaucoup plus proches des réalisations réelles des techniques d'apprentissages automatiques « connexionnistes ».

LES CONTROVERSES MÉDIATIQUES À PROPOS DE L'INTELLIGENCE ARTIFICIELLE

Une première caractéristique des débats autour de l'IA est que l'ensemble des technologies auxquelles ils font référence, même si elles ont pour certaines une longue et tumultueuse généalogie dans l'histoire de la discipline informatique, recouvre en fait une gamme de services et d'applications extrêmement floue. Si les termes d'IA et d'algorithme dominent aujourd'hui le débat, ils restent cependant très imprécis tant ils recouvrent des ensembles de dispositifs plus ou moins définis et complexes (robotique, algorithme, dialogue naturel, *machine learning*, apprentissage profond, système autonome, etc.). Le succès populaire du terme d'IA tient à sa manière de rassembler des gammes de technologies très différentes derrière un signifiant qui appelle presque immédiatement une interrogation ontologique sur la nature de l'intelligence et le statut de la vie artificielle (Katz, 2017). Afin de se donner une définition opératoire et agnostique de l'IA, nous privilégions dans cet article les termes d'agent calculateur ou de dispositif de calcul pour englober l'ensemble des techniques qui par le truchement de procédures automatisées produisent un résultat en sortie à partir de tout type de données. Cette définition englobante nous permet de ne pas limiter *a priori* ces entités pour suivre la manière dont les acteurs vont les définir et prêter une agentivité aux réseaux d'entités qui y sont associés (terminaux, codes, fonctions, données, développeurs, utilisateurs, etc.) dans des agencements sociotechniques spécifiques (Seaver, 2017).

Une deuxième caractéristique de ces débats est qu'ils traversent l'espace social à des échelles variées et en endossant des formes différentes. Ils font d'abord l'objet d'une prise en charge spécifique par les sphères de l'expertise et de la décision publique. Depuis quelques années, face au développement et aux promesses de l'IA et afin de réguler son utilisation, de nombreuses initiatives publiques et privées visant à fonder une éthique sont apparues (Benbouzid *et al.*, 2022). Ces initiatives consistent le plus souvent en la production de chartes ou de déclarations par des experts de différents domaines, définissant les grands principes éthiques que doivent respecter les développeurs et promoteurs de ces technologies. En 2020, selon le Global Inventory of AI Ethics Guidelines, on ne comptait pas moins de 160 documents de recommandations éthiques produits par des experts, chercheurs, entrepreneurs et décideurs publics (Algorithm Watch, 2020). Des méta-analyses du contenu de ces chartes font apparaître une assez grande convergence des principes éthiques (Jobin *et al.*, 2019 ; Mittelstadt *et al.*, 2016 ; Flejd *et al.*, 2020) comme le respect de la vie privée, l'*accountability*, la transparence, l'explicitabilité, ou la

fairness. Bien que très prisées, les limites de ces chartes éthiques apparaissent d'emblée. La sincérité de ces démarches peut être remise en cause (*ethics washing*) lorsqu'elles sont émises par certains acteurs privés qui promeuvent et développent des solutions basées sur l'IA (Floridi, 2019). D'autre part, ces réflexions éthiques *top-down*, abstraites et principielles sont souvent d'une faible utilité lorsqu'il s'agit d'apprécier concrètement les conséquences du déploiement des dispositifs de calcul dans nos sociétés. En comparant ces initiatives éthiques avec les formes de régulation qui existent historiquement dans d'autres domaines comme la médecine, il apparaît qu'elles souffrent de nombreuses lacunes qui entravent leur implémentation. D'une part, les secteurs de développement de l'IA sont trop variés et concernent un champ d'application trop large qui tend à engendrer des tensions entre les intérêts privés et publics. D'autre part, du fait de son développement récent, il n'existe pas encore de culture professionnelle, de normes locales traduisibles en bonnes pratiques ou d'instance de régulation coercitive qui permettraient de réguler les pratiques à partir de principes génériques. Une des pistes pour pallier les difficultés d'application des principes éthiques est de localiser les enjeux par domaine d'application afin de traduire en pratique les principes par le bas dans des contextes bien définis (Mittelstadt, 2019). C'est une des directions prises aujourd'hui par les régulateurs nationaux et internationaux pour élaborer un cadre réglementaire contraignant à l'égard des services utilisant des techniques d'IA (Bareis et Katzenbach, 2021).

Parallèlement aux discussions éthiques, le débat autour de l'IA a aussi de nombreuses résonances au sein du grand public en raison de la densité des représentations populaires de ces technologies fournies par les œuvres de science-fiction (Hayles, 1999). Cette connexion entre le développement du programme scientifique de l'IA et l'imaginaire fictionnel s'est notamment fixé au cinéma dans les années 1960 avec *2001, l'odyssée de l'espace* de Stanley Kubrick dont Marvin Minsky était le conseiller scientifique. Le succès de ces œuvres de fiction assure une connexion directe entre le développement des innovations technologiques de l'IA et le grand public (Recchia, 2020). Mais celui-ci est aussi un utilisateur de techniques de calcul développant, lui aussi, des points de vue critiques lorsque dans la vie quotidienne il rencontre des troubles dans les calculs (Meunier *et al.*, 2019). Les utilisateurs de ces technologies sont alors conduits à interroger la manière dont ils sont calculés en s'intéressant aux sources de données ou au fonctionnement des algorithmes. Enquêter sur ces représentations ordinaires du fonctionnement des IA présente alors un grand intérêt pour comprendre la manière dont les individus imputent des intentions aux calculs ou à ceux qui les programment (Bucher, 2016).

Cependant, entre les chartes éthiques des experts institutionnels et l'expression de critiques ordinaires par les utilisateurs, il importe d'analyser les médiations qui transportent et traduisent les enjeux vers le public. Cet article cherche à combler cet espace en conduisant une observation des controverses entourant le développement de l'IA à travers les médias d'information générale.

La troisième caractéristique de ces débats publics est qu'ils s'organisent systématiquement autour d'une polarité tranchée entre risques et promesses et qu'ils se nouent souvent autour d'*affaires* (Boltanski et Claverie, 2007) prenant forme à partir d'une situation singulière de troubles provoqués par le calcul. La mise en débat des technologies de l'IA s'appuie alors sur la mise en évidence actuelle ou potentielle d'un risque de biais, de discrimination, de censure ou de disqualification consécutive à l'entrée des agents calculatoires dans la société. Aussi est-il intéressant de s'intéresser au sein de l'immense production de rapports d'experts, de discours de vulgarisation ou de promotion de ces technologies, aux spécificités des discours critiques. C'est en effet au niveau des médias que la dimension critique, la mise en visibilité de controverses autour d'affaires, de troubles et d'inquiétudes à l'égard des changements technologiques est la plus apparente. Entre les représentations ordinaires et les rapports d'experts, les médias constituent le principal espace de mise en scène des promesses et des risques de l'IA. Pour les étudier, nous avons entrepris une démarche d'analyse de corpus systématique en utilisant des outils de traitement automatique de la langue. Nous nous proposons d'identifier la manière dont les médias cadrent les problèmes liés au déploiement de l'intelligence artificielle. Quels types de calculateurs sont mis en cause ? Comment sont-ils définis ? Quels types de troubles suscitent-ils ou sont-ils susceptibles de produire ? Et qui en sont les victimes ? Cette approche exploratoire conduite à partir d'un grand corpus d'articles permet de faire apparaître deux registres de discours, la prophétie et la critique, qui constituent les deux faces contrastées, mais partiellement interdépendantes, des représentations des nuisances de l'intelligence artificielle.

CARTOGRAPHIE DE LA CRITIQUE MÉDIATIQUE DE L'IA ET DES ALGORITHMES

La cartographie sémantique d'un large corpus d'articles de presse portant sur l'IA fait d'abord émerger deux types d'entités différentes. D'une part, les *algorithmes* qui façonnent nos environnements de calcul quotidiens sont associés à un discours critique portant sur les biais, les discriminations, la surveillance,

la censure et les phénomènes d'amplification dans la diffusion de contenus inappropriés. D'autre part, les *robots* mettent en scène des entités autonomes, incarnées et indépendantes qui sont associées à un discours prophétique alertant sur notre capacité à contrôler ces agents artificiels qui mettent en péril les capacités physiques et cognitives des humains et leur modèle de société. Avant d'analyser les spécificités de ces deux régimes de discours, il est nécessaire de détailler le corpus et les méthodes que nous avons mis en œuvre pour les faire apparaître.

Visibilité croissante du thème de l'IA et détection d'un discours critique

Pour cette recherche, nous avons extrait un large corpus de 29 342 articles de presse en anglais portant sur l'IA et les algorithmes sur une période de cinq années de 2015 à 2019². Avec la Chine, les États-Unis et le Royaume-Uni occupent la première place sur le marché de ces technologies. Selon un rapport de l'OCDE (2019) les États-Unis est le pays qui connaît les plus importants investissements entre 2011 et 2018 dans les technologies d'IA à l'échelle mondiale, le Royaume-Uni est quant à lui également leader des investissements dans ces technologies à l'échelle européenne. La presse de ces deux pays méritait à ce titre une investigation spécifique. Le premier constat que cette collecte fait apparaître est une forte augmentation du volume d'articles portant sur l'IA sur la période d'observation. Cette augmentation atteste de la place de plus en plus importante du thème des algorithmes et de l'IA dans l'espace médiatique ces dernières années. La visibilité croissante du thème de l'IA a été observée dans d'autres travaux à partir d'analyses statistiques portant sur différentes sources informationnelles³. Afin de contrôler la spécificité

2. Ceux-ci ont été extraits à partir de la requête "artificial intelligence" OR "AI" OR "algorithm*" OR "machine learning" OR "deep learning" OR "neural network*" – "amnesty international" – "weiwei" – "air india", réalisée sur Factiva sur un ensemble de 47 sources de presse nationale dite généraliste aux États-Unis et au Royaume-Uni (27 sources pour USA et 20 sources pour UK). Les termes "amnesty international" et "air india" sont des *stop words* qui ont été insérés dans la requête car souvent présents avec l'acronyme AI, de même pour le terme "weiwei" associé à l'artiste AI Weiwei, afin de limiter le bruit dans les résultats de la requête. Le corpus de départ ainsi obtenu, couvrant une période de 5 années (01/01/2015 au 31/12/2019), est constitué de 29 342 articles. La sélection des sources a été effectuée en extrayant les listes de sources de presse généraliste quotidienne nationale aux États-Unis et au Royaume-Uni sur Wikipédia et référencées par l'agrégateur Factiva (voir annexe 1 pour la liste complète des sources).

3. Tout d'abord dans la presse anglophone, à l'échelle mondiale, à partir d'analyses statistiques produites par le projet GDELT (Perrault *et al.*, 2019), s'observe une croissance du thème

de cette augmentation au sein de notre corpus, nous avons comparé avec d'autres requêtes effectuées sur la même période et des sources identiques, pour lesquelles on attend une distribution stable. L'observation du volume d'articles extraits à partir de notre requête montre une augmentation significative de 163 % entre 2015 et 2019⁴ des publications portant sur l'IA et les algorithmes alors que les articles consacrés à l'art, la culture, l'économie ou la technologie restent relativement stables (figure 1).

Figure 1. Évolution temporelle du thème de l'IA et des algorithmes sur la période 2015-2018 comparée à d'autres requêtes sur la même base d'articles de presse (Factiva)

Requête Factiva	2015	2016	2017	2018	2019	taux de variation en % sur la période
AI & Algorithm						163%
Art						-9%
Culture						2%
Economy						0%
Technology						14%

Source : auteurs.

de l'IA et des termes associés tels que *deep learning*, *machine learning* entre 2017 et 2019. Cette augmentation de la présence du terme de l'IA est aussi largement visible à partir des statistiques de requêtes effectuées sur le moteur de recherche Google aux USA et analysables via le service Google trends sur la période allant de 2015 à 2019. Enfin, le même constat est également fait d'une augmentation entre 2017 et 2019 de la fréquence d'apparition des termes *AI* et *machine learning* dans différents corpus de documents issus par exemple d'organisations gouvernementales aux USA, au Canada, et au Royaume-Uni, mais également dans les appels de fonds des entreprises américaines.

4. Les statistiques comparées des différentes requêtes intègrent les doublons présents dans la base Factiva.

En s'appuyant sur des outils d'annotations par apprentissage supervisé⁵, un sous-corpus d'articles produisant un discours critique des algorithmes et de l'IA a été constitué. À partir de l'annotation manuelle de 2000 titres d'articles de presse jugés « critiques », nous avons construit un modèle d'apprentissage automatique et réalisé plusieurs itérations incluant des phases de contrôle des résultats afin d'optimiser le modèle de détection des articles « critiques ». Tous les titres d'articles comportant des arguments ou éléments sémantiques négatifs portant explicitement sur une forme de calculateur (algorithme, IA, robot, etc.) ont été catégorisés comme critiques. À l'inverse ont été annotés comme non critiques ou ignorés les titres d'articles aux énoncés neutres, positifs, ambigus ou qui n'évoquent pas directement un calculateur. Par exemple les titres “Robots put jobs at risk” ou “Robot lawyers: how humans can fight back” ont été annotés comme critiques, alors que “AI to create more than 7m jobs”, “In the 2020s, artificial intelligence will transform the work of lawyers” ou “Need a lawyer? There’s an algorithm for that” ont été annotés non critiques. Étant donné le travail de *sourcing* manuel du corpus de départ et l'usage de méthode d'annotation par apprentissage supervisé, ce corpus ne prétend pas à une quelconque exhaustivité ou représentativité. Notre démarche permet en revanche d'appréhender avec la plus large diversité possible les thèmes associés à la critique de l'IA et des algorithmes dans la presse généraliste. Afin de s'assurer de la variété et de la précision des articles, différents modèles d'annotation ont été comparés et améliorés par itérations successives et un contrôle manuel systématique des résultats produits par ces modèles a été effectué⁶.

Le taux d'articles annotés comme critiques est de 7,1 % en moyenne sur l'ensemble du corpus et présente une assez forte stabilité durant la période (écart-type de +/-2,2 %). Si le volume d'articles portant sur l'IA et les algorithmes augmente significativement, le taux d'articles détectés par le modèle comme contenant un discours critique augmente simultanément durant la période. Des travaux (Brennen *et al.*, 2018) portant sur l'analyse de la couverture médiatique du thème de l'IA en Grande-Bretagne ont montré qu'une très large majorité des publications portaient sur les avancées technologiques dans le domaine et sur les potentiels développements que ces derniers permettent dans les secteurs du commerce et de l'industrie. Les questions éthiques et les problèmes

5. Nous avons utilisé le modèle de *machine learning* Fasttext : <https://fasttext.cc>, consulté le 2 mars 2022.

6. La méthodologie mise en œuvre est détaillée dans Crépel *et al.* (2021).

sociaux tels que les risques de discrimination associés au déploiement de ces technologies sont sous-représentés au sein de l'espace médiatique.

La figure 2 présente le réseau sémantique obtenu à partir des 2 091 articles annotés comme critiques. Elle repose sur une extraction de termes les plus représentés au sein des articles⁷. Ce réseau est constitué de 2 991 termes (ou suite de termes) reliés entre eux par 54 062 liens de cooccurrences. La spatialisation est basée sur une variante de l'algorithme Fruchterman Reingold, incluant le poids des cooccurrences entre les termes du réseau ; la position est optimisée par leur appartenance aux clusters détectés automatiquement par l'algorithme de Louvain. La couleur des nœuds est relative à chacun des clusters ainsi détectés et leur taille traduit leur fréquence au sein du corpus⁸.

Une structure par agents calculateurs et domaines d'application

À partir des termes les plus représentés au sein de chacun des clusters, on peut interpréter et nommer les thèmes qui structurent l'espace sémantique de ce corpus d'articles critiques de l'IA et des algorithmes. La cartographie fait apparaître 23 clusters thématiques de tailles variées qui représentent des types de calculateurs ou des domaines d'application différents⁹. Une première lecture du réseau consiste à décrire les thèmes les plus présents au sein du cor-

7. Cette extraction a été réalisée avec le logiciel Cortext à partir de la méthode *pigeonhole*, au niveau des phrases sur le texte complet de chaque article. Voir <https://www.cortext.net>, consulté le 3 février 2022.

8. La spatialisation a été produite avec la plateforme Cortext, puis visualisée avec le logiciel Gephi (<https://gephi.org>, consulté le 3 février 2022). Chaque nœud du réseau représente un terme (ou suite de termes) extrait dans le corps du texte des articles analysés. Les liens entre les nœuds apparaissent lorsque des termes sont cooccurents au niveau de la phrase dans le texte des articles. Les positions des nœuds du réseau dépendent de la connectivité qui existe entre eux. L'algorithme de spatialisation du réseau permet de rendre compte dans un espace à deux dimensions des proximités entre les nœuds en fonction des liens qui les unissent de sorte que des nœuds proches dans le réseau seront souvent utilisés ensemble dans les articles. En appliquant un algorithme de modularité sur le réseau, on est capable de détecter automatiquement des zones de relations plus denses entre certains nœuds, c'est-à-dire au sein desquelles les termes sont souvent cooccurents dans les articles. À partir de cette spatialisation et de la détection automatique des clusters, on interprète qualitativement les thèmes associés à chacune des zones du réseau sémantique.

9. Pour chaque article, en fonction des termes mobilisés, il lui est assigné un seul et unique cluster. De la même manière chaque terme est associé à un unique cluster même s'il peut entretenir des liens de cooccurrence avec d'autres termes appartenant à des clusters différents.

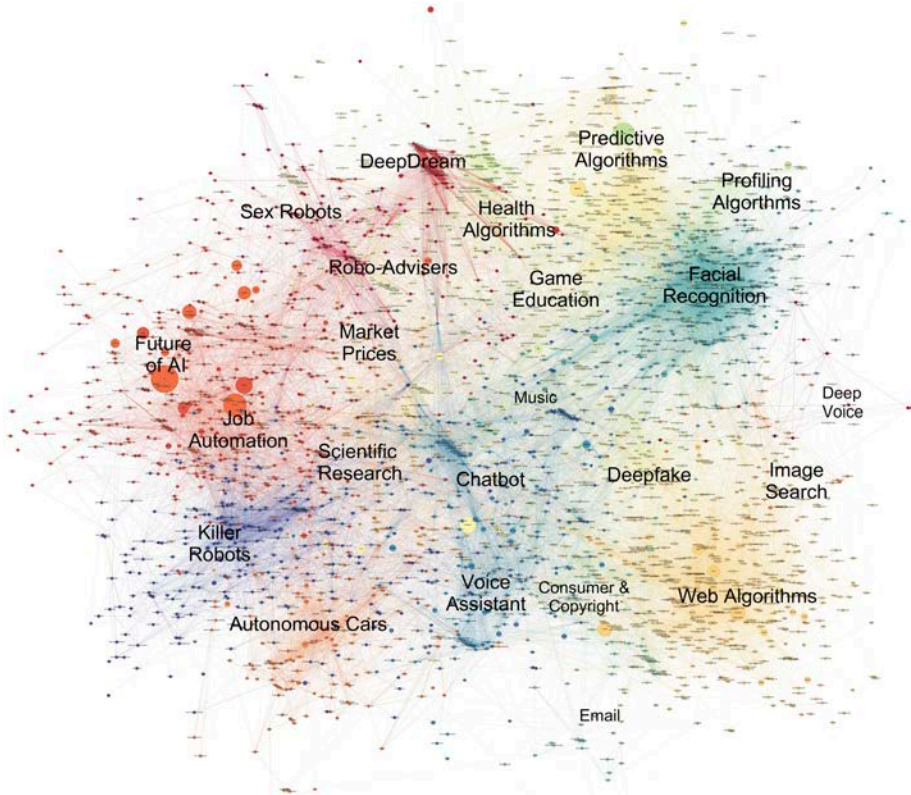
pus en nombre d'articles et d'en proposer une illustration succincte à partir d'exemples de titres d'articles prototypiques qui leurs sont associés.

Les deux plus importants clusters du graphe en nombre d'articles forment deux pôles opposés. Le plus important nommé "Web Algorithms", représentant 22 % des articles du corpus critique, comporte des articles qui traitent des troubles produits par les algorithmes du web tels que les techniques de classement du fil d'actualité de Facebook ("Facebook accused of censoring conservatives, report says" – *The Guardian* – 10/05/16), les recommandations des vidéos sur YouTube ("YouTube, the Great Radicalizer" – *The New York Times* – 11/03/18) ou encore des moteurs de recherche ("Google results claim that the Holocaust didn't happen, and company won't change it" – *Independent Online* – 15/12/16). À l'extrémité opposée du graphe, on trouve le second plus important cluster nommé "Future of AI" qui représente 18 % du corpus d'articles, évoquant les risques de l'émergence d'intelligences artificielles et de machines autonomes imitant ou surpassant les capacités humaines ("AI will create 'useless class' of human, predicts bestselling historian" – *The Guardian* – 20/05/16) ou menaçant l'humanité d'extinction ("End of Humanity? Artificial Intelligence could destroy us 'within decades' warns expert" – *express.co.uk* – 15/04/16).

Parmi les plus importants clusters, on trouve trois autres thèmes largement représentés dans le corpus. Le cluster "Job automation" (14 % des articles) contient des articles alertant sur les risques de mutation du marché de l'emploi face à la robotisation et au déploiement de l'IA dans le monde du travail ("Robots and AI are threatening close to a third of UK jobs, study reveals" – *Independent Online* – 24/03/17). Le cluster "Killer robots" (11 % des articles) est constitué d'articles sur les risques du déploiement de l'IA et de machines autonomes dans le cadre de conflits armés ("Top scientists call for ban on killer robots to prevent apocalyptic war" – *dailystar.co.uk* – 20/08/17). Enfin, le cluster « Facial recognition » qui représente 10 % des articles du corpus porte sur différents développements des technologies de reconnaissance faciale dans l'espace public, dans des logiciels ou sur les plateformes du Web ("China is now using facial recognition cameras to monitor Uighur Muslims across the country, report claims" – *Mail Online* – 15/04/19).

L'ensemble de ces cinq plus grands clusters thématiques domine largement l'espace médiatique car ces derniers ne constituent pas moins de 75,6 % du total des articles du corpus critique (1 581 articles) et représentent sur le graphe 46,4 % du réseau de termes extraits (1 389 termes).

Figure 2. Carte sémantique des thématiques des articles « critiques »
Explorer le graphe : <https://medialab.github.io/carnet-algopresse/#/publication/fr>, consulté le 3 mars 2022.



Source : auteurs.

Un second ensemble, bien que très hétérogène du point de vue des thématiques déployées et de leur position au sein du graphe, est constitué de 9 clusters qui représentent respectivement entre 1 % et 4 % du nombre d'articles du corpus. L'ensemble de ces thèmes représente 21,6 % des articles du corpus (452 articles) et 38 % des termes du graphe (1136 termes). Dans ce sous-ensemble se trouvent des clusters qui regroupent des types de calculateurs particuliers tels que les "Voice Assistant" ("Google's Nest microphone intensifies privacy concerns" – *The Washington Post* – 25/02/19), les "Autonomous Cars" ("Self-driving cars will need to be programmed to kill their owners, academics warn, and people will have to choose who will die" – *Independent Online* – 27/10/15), les "Sex Robots" ("Sex robots

could lead to population crisis as men opt for virtual girlfriends” – Mirror.co.uk – 27/01/19), les “Chatbot” (“Microsoft ‘deeply sorry’ after AI becomes ‘Hitler-loving sex robot’ – The Telegraph Online” – 26/03/16) ou encore les technologies de “Deepfake” (You thought fake news was bad? Deep fakes are where truth goes to die” – *The Guardian* – 2/11/18). On trouve ainsi des clusters autour des algorithmes de police et de justice prédictive, “Predictive Algorithms” (“Criminal justice software algorithm used across the US is biased against black inmates, study finds” – Independent Online – 27/06/16), d’identification et de ciblage de population avec les “Profiling Algorithms” (“Artificial intelligence can identify ‘gay faces’ from a picture, study claims” – Independent Online – 08/09/17), le domaine de la santé, “Health_Algorithms” (“Racial bias in a medical algorithm favors white patients over sicker black patients” – *The Washington Post* – 25/10/19) ou des jeux et de l’éducation, “Game and Education” (“Korean Go master quits the game because AI ‘cannot be defeated’” – CNN Wire – 28/11/19). Enfin le graphe se compose de 9 autres clusters de taille très réduite, entre 0,8 % et 0,1 % des articles pour chaque cluster (58 articles au total)¹⁰.

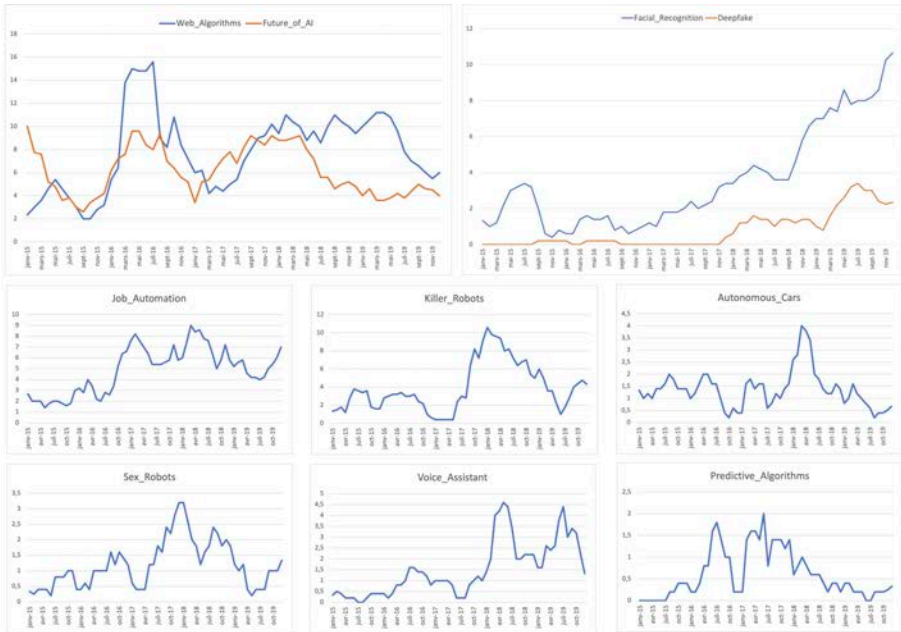
Stabilité et évolutions de l’agenda médiatique

L’analyse de la distribution temporelle du nombre d’articles critiques des dix plus importants clusters sur la période de cinq années fait apparaître trois trajectoires distinctes (figure 3). On observe une distribution assez stable sur la période de la répartition des articles des deux plus grands clusters structurants de la cartographie que sont les “Web Algorithms” et “Future of AI”. La distribution de ces deux plus grands ensembles montre qu’il existe une production régulière d’articles critiques sur les troubles liés aux algorithmes du web et aux craintes relatives aux développements futurs de l’intelligence artificielle et des robots. Ces thèmes concentrent le plus grand nombre d’articles et forment deux pôles structurants sur lesquels s’établit de façon récurrente un

10. Les termes extraits de ces articles représentent pourtant 15,6 % (466 termes au total) de la totalité des termes du graphe. Ces thématiques peuvent s’apparenter à des signaux faibles d’un discours critique sur les algorithmes et l’IA détectés par notre méthode d’extraction de termes mais pourtant sous représentés en volume d’articles. Ces clusters se centrent autour de thématiques très spécifiques, évoquant parfois des types de calculateurs (“Robot-Advisers”, “DeepDream Nightmares”, “Deep Voice”) mais le plus souvent des domaines d’application (“Scientific Research”, “Market and Prices”, “Image Search”, “Email”, “Music”, “Consumer and Copyright”).

discours critique dans les médias. Un second profil de distribution concerne les clusters “Facial recognition” et “Deepfake” qui connaissent une hausse importante du volume d’articles publiés sur l’année 2019 signalant la hausse soudaine de cette préoccupation. Enfin, le troisième profil de distribution montre une hausse du nombre d’articles qui se concentrent sur les années 2017 et 2018 et une baisse significative sur l’année 2019. Ce type de distribution concerne les clusters “Job Automation”, “Killer Robots”, “Autonomous Cars”, “Sex Robots”, “Voice Assistant”, “Predictive Algorithms”.

Figure 3. Distribution du nombre d’articles « critiques » des dix plus importants clusters sur la période couverte par le corpus

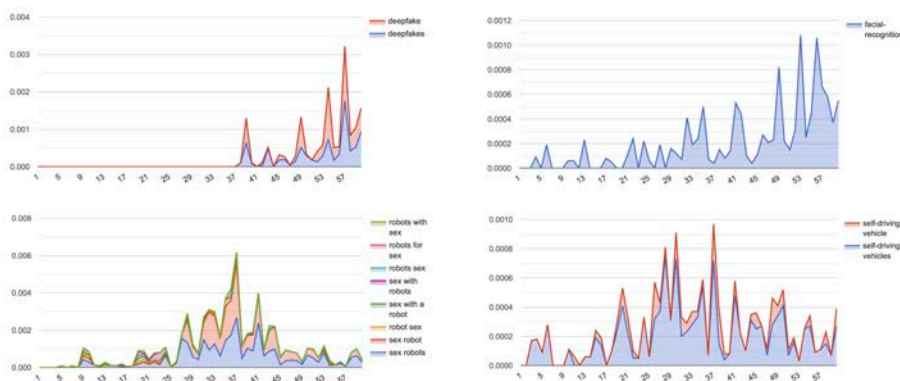


Source : auteurs.

Les différentes tendances exprimées par ces distributions doivent être interprétées avec prudence mais elles montrent la manière dont la presse, en focalisant ponctuellement son attention sur certaines technologies plutôt que d’autres, constitue un agenda médiatique suscitant un discours critique. La critique n’est pas consécutive mais immédiatement associée à l’entrée d’une nouvelle préoccupation dans les médias. Afin de vérifier cette hypothèse, nous avons observé les distributions de certains termes constitutifs de ces clusters

au sein du corpus d'articles étiquetés comme non critiques. Ces distributions montrent une répartition similaire à celle observée pour les clusters du corpus critique (figure 4). Ainsi les “deepfake” connaissent une fréquence d'apparition croissante au sein des articles non critiques entre 2018 et 2019. La même distribution est observable également pour le terme “facial recognition” qui croît progressivement jusqu'en 2019. L'ensemble des termes associés aux robots sexuels présents dans le corpus étiqueté comme non critique connaissent leur plus forte fréquence sur les années 2017 et 2018 avant de diminuer en 2019, on observe également le même phénomène de concentration de la fréquence d'apparition des termes relatifs aux voitures autonomes et aux assistants vocaux. Le léger reflux de ces thèmes montre qu'en phase initiale de cristallisation de la vague de l'IA les thèmes les plus à même de séduire les journalistes et les publics (voiture autonome ou robots sexuels) sont portés sur le devant de la scène avant d'être relativisés en raison du manque de réalité des performances des technologies annoncées¹¹. À la différence de ces thèmes suscitant des pics d'attention rapidement atténués, d'autres sous-

Figure 4. Distribution de la fréquence de termes au sein du corpus d'articles non critiques relatifs aux clusters “deepfake”, facial recognition”, “sex robots” et “autonomous vehicules” sur la période couverte par le corpus



Source : auteurs.

11. C'est le cas par exemple des voitures autonomes promises pour un avenir très proche en 2017 et dont les prévisions ont depuis été reportées dans le futur et dont l'« autonomie » s'appuie en réalité sur une multitude d'attachements socio-techniques à l'environnement (voir Tennant et Stilgoe, 2021).

thématiques récentes de l'IA comme la reconnaissance faciale et les *deep fakes* se sont, elles, installées comme des problèmes publics durables. Sous l'effet des inquiétudes publiques relatives à la qualité de l'information et à la généralisation de la surveillance, ces deux technologies sont désormais régulièrement incorporées aux discours publics sur les menaces de l'IA.

Topologie et axe de l'autonomisation

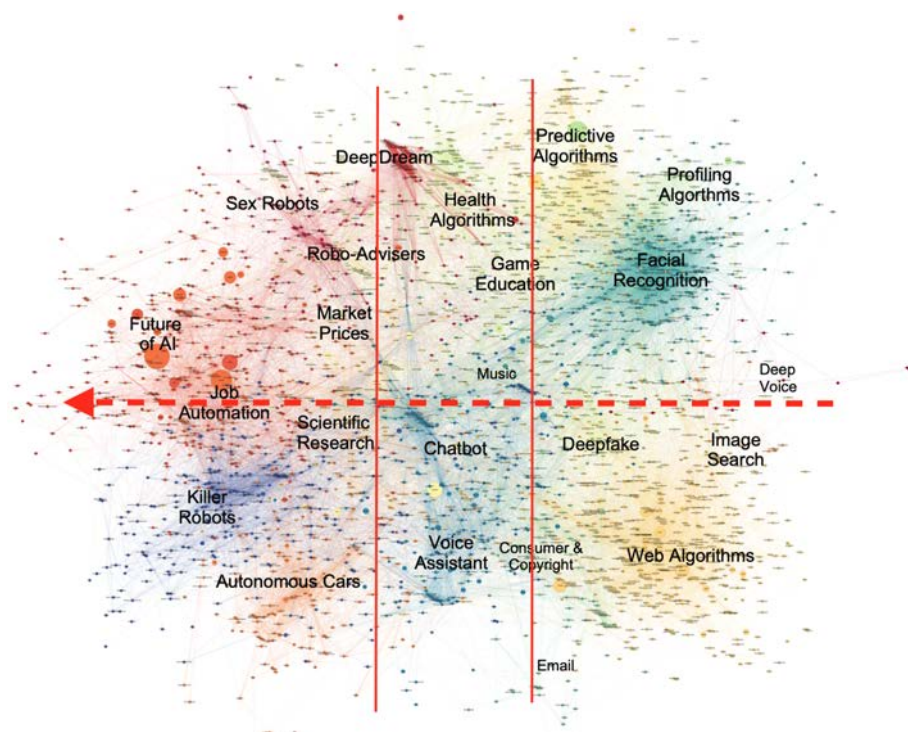
La dispersion des clusters de termes dans l'espace topologique laisse apparaître une opposition entre deux principaux types de calculateurs qui émergent lorsque l'on trace un axe horizontal allant de la droite à la gauche du graphe (figure 5). Cet axe peut s'interpréter comme un processus d'autonomisation progressive de l'artefact par rapport à son environnement. Il permet d'observer un déplacement entre des articles mettant en scène des techniques de calcul algorithmique incorporées dans l'environnement de l'utilisateur pour guider, orienter ou contrôler ses comportements, vers des articles qui se caractérisent par la personnification de l'IA dans une entité incarnée et autonome à l'égard des humains et de l'environnement. La première zone à droite du graphe concerne exclusivement des algorithmes, qu'ils s'agissent des algorithmes du Web comme le fil d'actualité de Facebook, la section *trending topic* de Twitter, les algorithmes de recherche d'images et de sites web de Google. Si la majorité de ces dispositifs sont ancrés dans les usages quotidiens, d'autres représentent des technologies plus émergentes qui font l'objet de développements et d'expérimentations plus localisées et plus régulées tels les algorithmes de police ou de justice prédictive, de détection des images dans les systèmes de reconnaissance faciale ou les *deepfakes*. Ils n'en restent pas moins des services intégrés dans l'environnement de plus en plus informationnel et « intelligent » de l'utilisateur.

Dans la partie la plus à gauche du graphe, les clusters se concentrent autour d'agents calculateurs qui, à l'inverse, s'incarnent sous forme de machines ou de robots autonomes. Les entités qui peuplent cette zone sont personnifiées dans des dispositifs tangibles séparés de nos équipements numériques habituels que sont nos ordinateurs et nos smartphones. Ces dispositifs, en plus d'être incarnés physiquement sont équipés d'une capacité d'agir autonome via de systèmes de calcul basés sur l'IA. En simulant à la fois le corps et les capacités cognitives des humains, ils sont en mesure de produire certaines actions sans intervention humaine dans différents domaines telles que le transport avec les voitures autonomes, la défense avec les « robots tueurs »

et les drones, le travail avec l'automatisation des tâches professionnelles ou encore les relations physiques avec les robots sexuels.

Entre ces deux pôles existe une zone intermédiaire constituée de dispositifs à mi-chemin entre les machines autonomes et les algorithmes de nos environnements numériques quotidiens. On y trouve par exemple les assistants vocaux pouvant être embarqués dans les smartphones ou dans des dispositifs du type enceinte connectée, les chatbots du Web, les robo-advisors qui sont des algorithmes effectuant des opérations dans le domaine de la finance ou encore des technologies de *computer vision* comme le dispositif Deep Dream de Google. Ces technologies bien qu'assez accessibles au grand public sont souvent encore émergentes. Elles sont en capacité de gérer un certain nombre de tâches parfois basiques de manière autonome mais ne sont pas toujours incarnées dans des artefacts indépendants ; le plus souvent, elles sont des services accessibles via nos ordinateurs et smartphones. Un autre exemple de cette

Figure 5. L'axe de l'autonomisation de l'IA



Source : auteurs.

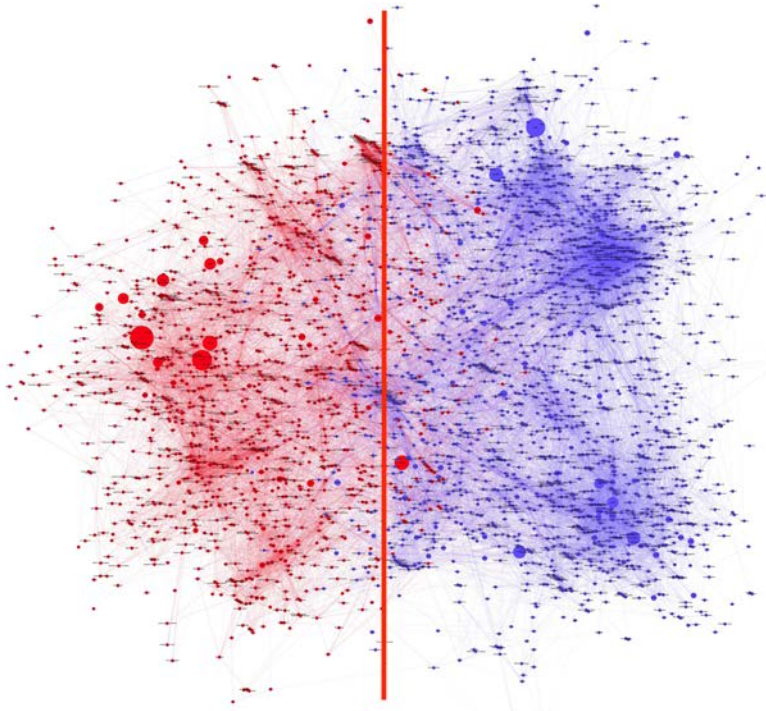
zone intermédiaire est la présence du cluster nommé “Health Algorithms” qui regroupe un ensemble d’articles portant à la fois sur des applications de détection automatisée des cycles menstruels dans des applications mobiles grand public mais également l’utilisation de méthodes de *machine learning* innovantes dans le domaine de la radiologie médicale.

La tension que fait apparaître cet axe rejoue et prolonge un clivage profondément enraciné dans l’histoire et l’imaginaire de l’informatique. On peut en faire remonter l’origine aux conceptions antagonistes de l’articulation entre informatique et société de John McCarthy et Doug Engelbart, tous deux à la tête des deux principaux laboratoires d’informatique de Stanford à la fin des années 1960 (Markoff, 2015). Pour John McCarthy, qui a forgé en 1956 avec le groupe du workshop de Dartmouth le terme d’intelligence artificielle, le projet scientifique de l’IA est de conférer une intelligence aux machines en leur donnant des capacités cognitives susceptibles de reproduire le raisonnement humain jusque dans la capacité de se donner à elles-mêmes des objectifs. Cette vision de l’IA qui a nourri le projet symbolique de la “Good Old-Fashioned Artificial intelligence”, et initié le développement de la robotique, cherche à conférer une intelligence propre à la machine afin de la rendre autonome (Haugeland, 1985 ; McCorduck, 2004). À l’inverse, la vision proposée par Doug Engelbart d’augmenter l’intelligence des humains grâce à l’informatique conçoit d’abord la technologie comme une manière d’équiper l’environnement d’outils de communication et de calcul qui potentialisent les capacités des humains. Dans cette vision, l’intelligence se trouve moins dans les machines que dans le couple qu’elles forment avec leur environnement sociotechnique (Bardini, 2000). Ces deux visions constituent une polarité toujours vivace dans l’histoire des technologies de l’informatique et singulièrement de l’intelligence artificielle. Elles ont donné lieu au développement de gammes de technologies et de services différents dont témoigne très explicitement la cartographie des thématiques que nous venons de détailler. D’un côté, à la suite du projet de machine intelligente, les projets de robots, de voiture autonome ou d’interfaces de dialogue en langage naturel cherchent à réaliser le projet scientifique de John McCarthy. De l’autre, les entreprises visant à équiper l’environnement de dispositifs destinés à guider, orienter et aider à la décision les humains prolongent et perpétuent le projet d’un couplage de plus en plus étroit entre les humains et les machines initié par Joseph Licklider, Doug Engelbart et beaucoup de ceux qui sont à l’origine du développement d’internet (Grudin, 2009). Dans la cartographie des enjeux contemporains de l’IA que nous venons de constituer, ces services se retrouvent dans le pôle Algorithmes.

MONDE DES ROBOTS, MONDE DES ALGOS

La séparation entre les mondes des robots et des algorithmes que nous venons de faire apparaître dans la topologie des termes issus des articles de presse permet maintenant d'interroger la manière dont les thématiques distribuées sur l'axe de l'autonomisation de l'IA font apparaître deux registres discursifs distincts. Afin de réaliser une analyse comparative de ces deux pôles, le réseau a été découpé en deux zones constituées de deux sous-ensembles comparables en termes de volume d'articles (figure 6). À gauche de la cartographie,

Figure 6. Séparation du réseau de termes et du corpus d'articles associés en deux sous-ensembles équivalents, à gauche intitulé "robots" et à droite intitulé "algorithmes"



Source : auteurs.

l'espace intitulé "Robots" compte 53 % des articles du corpus (9 clusters et 1 790 termes). L'espace intitulé "Algorithmes", à droite de la carte, représente 47 % des articles (14 clusters et 1 201 termes). Ce découpage en deux sous-ensembles équivalents permet d'explorer les caractéristiques de ces deux registres discursifs.

Pour les mettre en évidence au moyen de méthodes de traitement automatique de la langue, nous avons identifié systématiquement la manière dont sont désignées les entités techniques, la société qui les abrite et la façon dont les premiers ont des effets sur la seconde. Tous ces discours critiques cherchent à qualifier la manière dont des *Agents techniques* produisent des *Victimes* en exerçant un type particulier d'opération ; ces dernières sont constitutives d'un ensemble de troubles révélant autant d'*Enjeux* produits par le déploiement des techniques d'IA¹².

Pour différencier ces actants (*Agents techniques*, *Victimes*, *Enjeux*) au sein du graphe, la plupart des termes ne souffrant pas trop d'ambiguïté (66 % des termes du graphe) ont été annotés manuellement en dix catégories distinctes¹³. Notre analyse consiste à comparer les catégories qui apparaissent particulièrement pertinentes pour différencier les deux espaces¹⁴. La catégorie "Technical" renvoie à ce que nous avons désigné comme des calculateurs. Elle

12. Dans un travail sur les algorithmes utilisés dans les services urbains, nous avons proposé une approche plus qualitative de ce modèle actantiel (Cardon et Crépel, 2019).

13. "Technical" (15 % des termes) : Calculateur, Machine, Fonctionnalité, Devices, Service Technique ou produit de startup difficiles à différencier de l'entité économique ; "Company" (5 % des termes) : Entreprises, Start-up, Domaine d'entreprise ; "Fiction" (1 % des termes) : personnages, films, jeux vidéo ; "Person" (6 % des termes) : Personnalités politique, médiatique, recherche, noms propres ; "People" (8 % des termes) : personne ou groupe de personnes non nommées ; "Institution" (4 % des termes) : Agence gouvernementale, ONG, Association, Université, Textes juridiques ; "Locality" (1 % des termes) : Lieux ; "Data_inout" (10 % des termes) : Données, traces ; "Topic" (10 % des termes) : Domaine d'application ; "Issues" (7 % des termes) : Enjeux et problèmes.

14. Il est important de noter que la position de certaines entités au sein de chacun de deux grands clusters Robots et Algorithmes peut parfois sembler incohérente avec l'analyse comparative qui en est proposée. Ces positionnements peuvent en partie s'expliquer par la méthode, basée sur une spatialisation du réseau sémantique, qui tend à donner une position centrale à certains nœuds qui ont des connexions avec d'autres nœuds positionnés à différents endroits du réseau éloignés les uns des autres. Le choix de découper le réseau en deux grands sous-ensembles peut ainsi faire basculer certains nœuds occupant une position centrale dans le réseau (à la frontière des deux sous-ensembles). Pour autant, les incohérences observées ne concernent que quelques termes pour lesquels on attendrait parfois un positionnement différent

rassemble les termes relatifs aux agents techniques tels que les algorithmes, les machines, les fonctionnalités ou services techniques, les terminaux mais également les produits ou services issus le plus souvent de startup qui sont parfois difficiles à différencier de l'entité économique qui les a produits. La seconde catégorie relative aux agents humains, nommée "People", concerne tous les termes désignant des *personnes* ou *groupes de personnes* non nommées (la catégorie "Person" rassemble les noms propres) qui sont présentés le plus souvent comme les victimes des calculateurs. Enfin la troisième catégorie, nommée "Issues", permet d'identifier les termes qui désignent des *enjeux*, des difficultés ou des problèmes extraits au sein des articles.

Qui sont les Agents ?

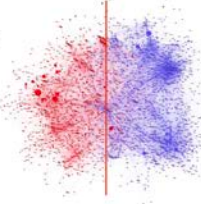
Les *Agents techniques* de nos deux sous-espaces sont des entités aux propriétés très différentes (figure 7). Ce qui caractérise la zone "Robots" est le très haut degré d'intelligence et d'autonomie des IA, ainsi que leur capacité à être incarné sous la forme de machines dotées d'une sorte de personnalité propre. Le terme de *robot* domine cet espace et que ce soit pour la conduite automobile, le sexe, la guerre ou la production industrielle, ces *Agents* sont dotés de capacité d'initiatives qui leur sont propres. *Robots, machines, computer, vehicles, weapons, drones* ou *dolls* sont omniprésents dans les articles. Certaines de ces entités techniques se distinguent car elles renvoient à une dénomination plus abstraite et générique des calculateurs désignés en tant que *system, artificial intelligence* ou encore comme *automation* ou *model*. Les caractéristiques données à ces entités contribuent à détacher leur capacité de choisir, de décider et d'agir du système socio-technique qui les a produits ou de l'environnement dans lequel elles se déploient. Si quelques entreprises apparaissent dans cette zone, comme Tesla, Uber, IBM, Nvidia ou Boston Dynamic, la plupart du temps les services évoqués n'ont pas d'existence très concrète ni de réalisation industrielle spécifique (figure 8). Les articles évoquent des projets industriels de la *Silicon Valley*, parlent de *firm* ou de *company* sans véritablement inscrire ces artefacts dans le paysage économique.

au sein des deux sous-ensembles, sans pour autant affaiblir l'analyse des termes dominants dans chaque espace sémantique.

Figure 7. Extrait des termes les plus cooccurents au sein du graphe relatifs aux entités techniques catégorisée “Technical” dans les deux espaces sémantiques “Robots” et “Algorithms”

Robots

AI, robots, machines, system, artificial intelligence, computer, tech, software, cars, automation, model, weapons, drones, vehicles, killer robots, sex robots, autonomous weapons, self-driving cars, neural networks, cloud, Watson, doll, weapons systems, superintelligence, sex dolls, humanoid, autonomous cars, humanoid robot, Sophia, Autopilot, robot arm



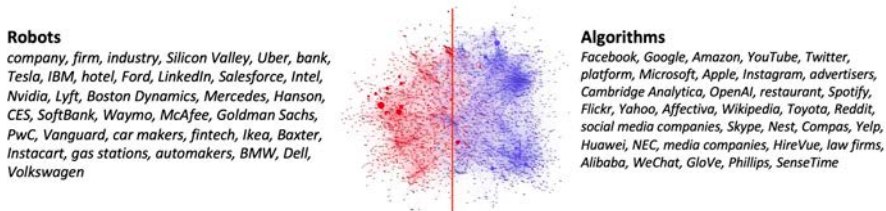
Algorithms

algorithm, app, devices, program, assistant, facial recognition, bot, phone, feature, Alexa, deepfakes, social network, Siri, chatbot, search engine, Echo, Tay, Android, Duplex, trending topics, Google Assistant, chat, iPhone, smart speakers, recommendation algorithm, surveillance technology, PredPol, Google Images, Facebook Messenger, image search, criminal justice algorithms

Source : auteurs.

En revanche, les *Agents* de la zone “Algorithmes” n’ont pas cette autonomie et sont plutôt définis comme des parties ou des éléments de systèmes de calcul distribués dans des environnements numériques. Ils ne présentent pas de caractéristiques anthropomorphes et relèvent plus de dispositifs d’augmentation de l’intelligence des environnements, contribuant à renforcer ou limiter les capacités d’actions des acteurs humains. Même sous des dénominations abstraites, les *Agents techniques* ne constituent pas des objets propres aux contours bien délimités par rapport à l’espace dans lequel ils agissent (*algorithm, devices, program* ou *phone*) (figure 7). La majorité d’entre eux renvoie à des dispositifs spécialisés comme *facial recognition, deepfake, social network, chatbot, criminal justice algorithm* et de manière encore plus spécifique à des fonctionnalités de certains services embarqués dans les terminaux mobiles ou sur le web, parfois associés à des marques, tels que *Siri, Search engine, Trending topics, Google assistant, Iphone, recommendation algorithm, Facebook messenger, Google images, image search*. Ces qualifications n’isolent pas les IA mais les incorporent dans un environnement sociotechnique dans lequel le calcul joue un rôle de plus en plus intense pour assister, trouver de l’information, guider ou prévenir un comportement. Cette partie du graphe concentre la plupart des mentions des grandes entreprises de services numériques cités dans les articles où l’on retrouve par ordre d’apparition Facebook, Google, Amazon, YouTube, Twitter et Microsoft (figure 8). L’IA des Algorithmes est celle des prédictions, des interactions avec l’utilisateur et du traitement de l’information à très large échelle qui s’est désormais introduite dans la plupart des fonctionnalités des grands services du web.

Figure 8. Extrait des termes les plus cooccurents au sein du graphe relatif aux entités économiques catégorisées “Company” dans les deux espaces sémantiques “Robots” et “Algorithms”



Source : auteurs.

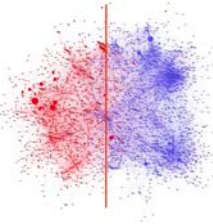
La catégorie “People” à partir de laquelle ont été annotés les termes désignant des agents humains, personnes ou groupes de personnes non nommées ou n’ayant pas d’existence institutionnelle, permet d’observer la manière dont est composé le monde social dans lequel les IA ont été introduites (figure 9). Dans les articles critiques de notre corpus, elle correspond presque toujours à des personnes désignées comme les victimes des opérations des Agents calculateurs. Une nouvelle différence apparaît alors très clairement entre les victimes des Robots et des Algorithmes. Du côté de l’espace “Robots”, on observe parmi les termes les plus fréquents une forte présence de références génériques à l’humanité, telles que *humans, humankind, human civilisation, human driver, human supervisor*. Les Robots n’ont qu’une seule cible : les humains. Cette procédure d’abstraction contribue à la dramatisation de la menace en unifiant les victimes dans un même destin et en les invitant toutes à une même réaction : la peur. D’autres termes assez génériques sont également présents dans cet espace mais relèvent d’une définition plus centrée sur des domaines dans lesquels les technologies robotiques sont amenées à se déployer, tels que la finance, le marché du travail, la défense et les transports. On retrouve alors, avec des fréquences moindres, des termes tels que *workers, customers, employees, drivers, retailers, soldiers, passengers, brokers, traders, farmers*. L’espace associé aux “Algorithmes” est quant à lui peuplé de termes beaucoup moins génériques que dans l’espace sémantique des “Robots”. Les agents humains sont ici désignés comme un ensemble de personnes ou de groupes sociaux qui occupent des rôles et des positions différentes au sein de la société – *i. e.* à la différence de l’abstraction des *humains*. On retrouve ainsi, parmi les termes fréquents, des entités faisant référence aux plateformes numériques *users, accounts, Facebook users, YouTube users*. Les termes qui désignent ces agents permettent dans la grande majorité de les

qualifier en fonction de propriétés comme l'âge (*children, child, kids, parents*), le genre (*women, men*), l'appartenance ethnique (*black people, black patients, african-americans*), l'orientation politique (*white supremacists, black defendants, illuminati*) ou sexuelle (*gays, lesbian, trans people*). On peut interpréter cette spécialisation des termes comme un mode de différenciation des qualités déposées sur l'identité des personnes par les caractéristiques de l'environnement sociotechnique auquel elles sont articulées.

Figure 9. Extrait des termes les plus cooccurents au sein du graphe relatif aux entités humaines catégorisées “People” dans les deux espaces sémantiques “Robots” et “Algorithms”

Robots

humans, humankind, human civilisation, human drivers, human supervisors, workers, customers, employees, consumers, drivers, investors, competitors, retailers, soldiers, contractors, passengers, Google employees, brokers, lenders, Uber drivers, traders, farmers, researchers and engineers



Algorithms

users, person, account, women, children, men, conservatives, players, students, patients, parents, child, profile, candidates, kids, gays, Facebook users, black people, white supremacists, innocent people, African-Americans, black patients, black defendants, lesbian, illuminati, white teenagers, trans people, YouTube users

Source : auteurs.

Quels sont les enjeux ?

La catégorie intitulée “Issues” permet une lecture comparée des termes relatifs aux troubles et aux enjeux qui sont associés aux agents techniques sur lesquels porte la critique (figure 10). Elle mérite d’être comparée à une analyse spécifique des verbes des articles comparant leur fréquence relative au sein des deux sous-ensembles du réseau (figure 11)¹⁵. Du côté de l’espace sémantique “Robots”, l’affrontement entre les humains et les IA est constamment

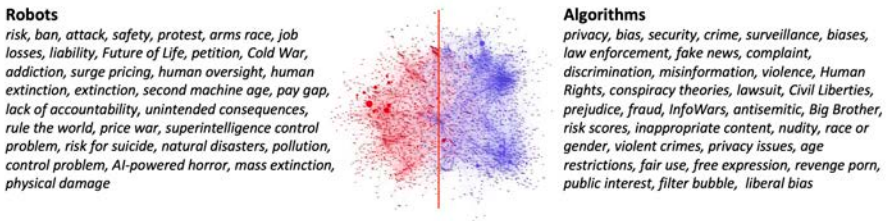
15. Nous avons réalisé une extraction spécifique des 1 000 verbes interprétables des articles par la méthode *pigeonhole* au niveau de la phrase au sein de chaque document via le logiciel Cortext en procédant à une sélection manuelle des verbes les plus interprétables parmi la liste initiale. Calcul de la fréquence relative d’apparition d’un verbe au sein des deux sous-ensembles d’articles. Plus précisément, pour une entité i , nous mesurons le score suivant : $s(i) = \log(p(i|Algos)/p(i|Algos))$. $s(i)$ est positif lorsque la fréquence relative de l’entité i est utilisée plus souvent dans le sous-corpus Algorithms. Inversement, les valeurs négatives correspondent à des entités qui sont concentrées dans le sous-corpus Robots. La hauteur des barres mesure l’importance de l’écart.

mis en scène comme une confrontation ultime mettant en jeu le devenir de l'humanité : « AI could become an immortal dictator from which we would never escape » (*Washington Post*, 6/4/18) ; « But our fate at the hand of clever cloggs robots may in fact be worse – to summon a class of eternally useless human beings » (*The Guardian*, 20/5/16). La présence renforcée de termes tels que *attack, safety, arms race, Cold war, human extinction, extinction, natural disasters, AI-powered horror, mass extinction, physical damage* renvoient le plus souvent au registre de la guerre ou de destruction à l'échelle planétaire. Les verbes les plus représentés dans les articles qui composent l'espace sémantique "Robots" confortent cette analyse des Enjeux. On observe une sur-représentation de verbes exprimant la destruction (*doom, destroy, eradicate, kill, eliminate*) et la domination des machines sur l'homme (*enslave, dominate*), mais également le dépassement ou le remplacement des hommes par ces agents techniques (*overtake, surpass, replace, defeat*). Ces menaces font émerger des enjeux de contrôle de ces technologies autonomes, tels que *ban, petition, human oversight, lack of accountability, superintelligence control problem, control problem*. D'autres termes relèvent de troubles assez visibles dans cet espace sémantique associés au remplacement, au dépassement et à une perte de contrôle de l'homme par les machines, notamment dans le domaine de l'économie, avec des termes tels que *job losses, surge pricing, pay gap*. D'autres verbes davantage représentés dans le sous-ensemble "Robots" renvoient à des notions de transformation et de changement (*reshape, transform, disrupt*) ou aux capacités de ces agents techniques à imiter ou simuler les comportements humains (*resemble, simulate, reproduce, mimic*).

Les termes relatifs aux troubles et enjeux constitutifs du sous-ensemble nommé "Algorithmes" laissent apparaître un autre registre de la critique qui se réfère à des questions de justice et d'injustice, ainsi qu'à des questions légales. En effet, parmi les termes fréquents, on trouve dans cette partie du graphe un champ sémantique constitué de références juridiques tels que *crime, law enforcement, Human Rights, lawsuit, Civil Liberties, prejudice, fraud, public interest* (figure 10). Les troubles produits par les agents techniques dénoncés au sein des articles associés à cet espace concernent les discriminations (*bias, biases, discrimination, antisemitic, race or gender, fair use, risk score, liberal bias*), les enjeux de protection de la vie privée (*privacy, surveillance, Big Brother, privacy issues*), les difficultés de filtrage ou d'exposition à des contenus inappropriés (*violence, inappropriate content, nudity, age restriction, violent crime*) et frauduleux (*fake news, misinformation, conspiracy theories, revenge porn, filter bubble*), ou encore de censure et de liberté d'expression (*free expression*). Les verbes associés au sous-ensemble

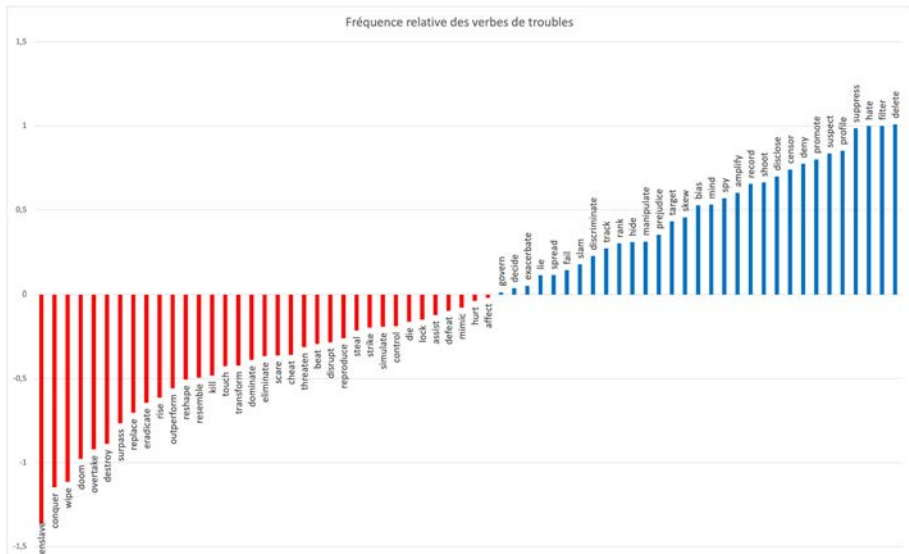
“Algorithmes” confortent là aussi les caractéristiques du vocabulaire des Enjeux (figure 11). En effet, les plus représentés renvoient aux questions de filtrage et de censure de l’information (*filter; delete, supress, censor*), aux problèmes de surveillance et de privacy (*profile, suspect, spy, target, track*), à la dénonciation des formes de discrimination (*bias, discriminate*) ou encore

Figure 10. Extrait des termes les plus cooccurrents au sein du graphe relatif aux entités troubles catégorisées “Issues” dans les deux espaces sémantiques “Robots” et “Algorithmes”



Source : auteurs.

Figure 11. Distribution des fréquences relatives des verbes de troubles entre les corpus d’articles associés aux “Robots” et “Algorithmes”



Source : auteurs.

aux phénomènes de propagation et d'amplification de contenus frauduleux (*promote, amplify, spread*). Toutes ces qualifications ne mettent plus en jeu l'humanité, mais font apparaître des souffrances, des discriminations et des inégalités au sein de la société.

Quelle est la temporalité ?

Cette distinction entre l'énonciation de la peur et de la critique s'exprime encore à travers les marqueurs temporels présents dans les articles¹⁶. Les extrapolations de la prophétie d'une humanité dominée par les robots procèdent à un allongement des marqueurs temporels contribuant à brouiller la frontière entre fiction et réalité pour faire advenir un futur apocalyptique. Les marqueurs de temps dans le sous-espace Robots relèvent d'une temporalité de long terme, qu'ils fassent référence au passé ou au futur (figure 12). Une grande majorité d'entre eux se structurent autour d'expressions qui utilisent la figure *the next* associée à des temporalités qui se comptent souvent en dizaines ou centaines d'années : *the next 30 years, the next 50 years, the next 100 years, the next 1,000 to 10,000 years*. D'autre part, les énoncés qui ont les scores les plus importants dans cet espace ont comme particularité de ne pas définir précisément les échéances auxquelles ils renvoient (*the next few decades, coming decades, the next century, this century*). Cet élargissement flou de l'espace temporel à l'échelle d'une histoire totalisante de l'humanité, à la manière des fresques à succès de Yuval Noah Harari (2017), se remarque aussi par la présence de marqueurs temporels qui ne font pas référence au futur mais au passé ; ceux-ci procèdent de la même manière à une projection très lointaine (*150 years ago, the past decade, a decade ago*). Cette indétermination temporelle est inséparable du fait que les technologies du sous-espace Robots sont encore, et toujours, en cours de développement au sein de laboratoires de recherche ou de start-up. Elles sont les vecteurs de l'incessante pro-

16. À partir de notre corpus, nous avons extrait 1000 entités nommées qui sont des marqueurs temporels présents au sein des articles – entités nommées intitulé "Date" dans le modèle d'extraction de terme de Spacy réalisée via le logiciel Cortext et de la même manière que pour les verbes, nous avons calculé de manière comparative leur fréquence relative sur l'ensemble des articles constituant les deux sous-ensembles "Robots" et "Algorithmes". La liste des entités extraites au départ a été filtrée manuellement afin de ne conserver que les marqueurs temporels interprétables en dehors de leur contexte d'énonciation. Nous avons supprimé les entités produisant des ambiguïtés ou qui ne sont pas interprétables hors de leur contexte d'énonciation comme les dates spécifiques ou chiffrées, les durées (ex. *several years, winter*), les fréquences (ex. : *everyday, weekly*) et les événements spécifique (ex. : *christmas*).

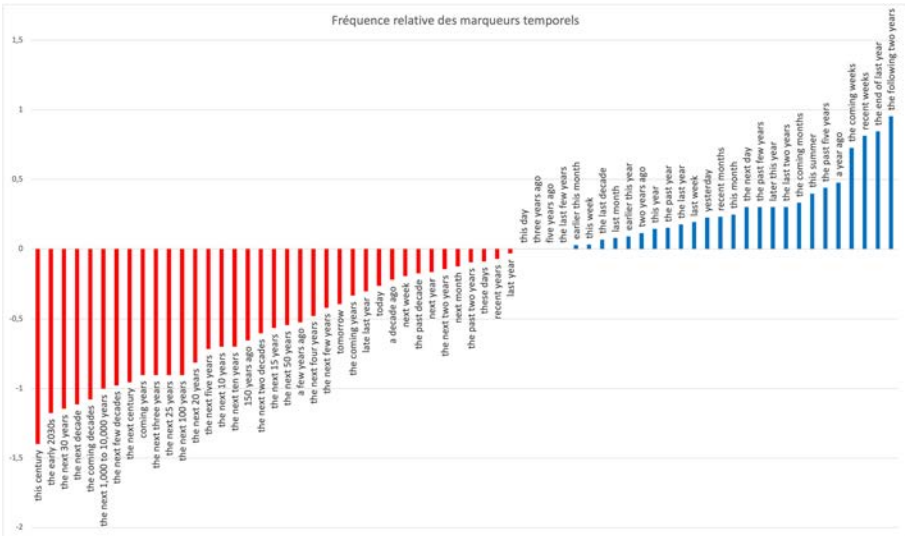
messe d'innovations de rupture qui dans le domaine du travail, de la guerre ou de sexe nourrit les représentations d'un monde sans travailleur, sans soldat et sans partenaire sexuel. Les clusters "Scientific research" et "Future of AI" enferment tous deux les recherches plus fondamentales dans les domaines de la robotique et de l'intelligence artificielle générale qui continuent de nourrir les promesses d'un raisonnement symbolique et autonome des machines (Marcus et Davis, 2019). Cependant, comme la parousie chrétienne, bien que toujours imminente, les repères temporels permettant de donner un calendrier à l'avènement d'une IA autonome sont particulièrement instables. Cette indétermination oblige la prophétie à constamment réviser son horizon, retouchant sans scrupule le calendrier de la promesse pour faire face au manque de réalité des prouesses attendues¹⁷.

C'est en revanche le présent proche qui domine les ancrages temporels au sein du sous-ensemble d'articles des Algorithmes. Les marqueurs temporels ayant les fréquences relatives les plus élevées se distinguent des premiers, car ils renvoient majoritairement au présent (*this week, this month, this summer*), à un passé très proche souvent exprimé en jour, semaine ou mois (*yesterday, recent weeks, recent months, the end of last year, a year ago*), mais également à un futur très proche en comparaison du premier sous-ensemble (*the next day, the coming weeks, the coming months, the following two years*). Les risques dénoncés dans les articles de ce sous-espace portent sur des systèmes techniques largement déployés dans l'environnement technologique quotidien des utilisateurs, tels que les algorithmes du Web, l'automatisation des emails, les algorithmes de moteurs de recherche d'image ou encore la synthèse vocale. D'autres technologies commencent à se diffuser auprès du grand public tels que les assistants vocaux, les "Deepfake" via des applications mobiles telles que Reface ou Faceswap. Enfin d'autres dispositifs ne sont pas encore démocratisés mais font pour la plupart l'objet d'expérimentations plus ou moins importantes par des institutions et des sociétés telles que la reconnaissance faciale, les algorithmes prédictifs dans le domaine de la police et de la justice qui connaissent actuellement des développements locaux, à l'échelle d'une institution ou d'un territoire. La forme prise par le discours critique n'a pas besoin de déporter dans un horizon temporel lointain les inégalités, les discriminations ou les effets de contrôle des techniques de calcul. Ceux-ci constituent un ensemble de nuisances présentes et actuelles.

17. Voir dans le cas de la singularité de Ray Kurzweil : Ganascia (2017).

Comme l’ont montré les travaux sur les formes argumentatives des énonciations du futur technologique (Chateauraynaud, 2013, 2014 ; Chateauraynaud et Debaz, 2019), les échelles temporelles jouent un rôle décisif dans la délimitation d’un registre de discours suffisamment ancré dans les réalités techniques et institutionnelles du moment pour favoriser une critique sociale, susceptible de désigner ses ennemis et d’imaginer des solutions politiques pour limiter leurs nuisances, et une prophétie technologique qui projette dans un avenir indéfini une menace globale dont il est beaucoup plus difficile de définir les responsables et qui, d’une certaine manière, se soustrait à la politique. Les épreuves de cohérence et de crédibilité dans les controverses nécessitent de définir des échelles temporelles plus resserrées qui vont venir modifier les jugements mais aussi les logiques d’action qui vont y être associées. En mobilisant en revanche des marqueurs temporels qui articulent des références au passé et une projection à long terme les acteurs établissent des prises sur le futur pour donner consistance à la menace, développer un discours d’urgence et enjoindre la société à mettre en place des mesures pour faire face à la situation sans pourtant être véritablement en mesure de les préciser tant ce futur indésirable apparaît inéluctable. Empruntant beaucoup de schèmes aux discours religieux, une rhétorique de la peur fondée sur une sorte

Figure 12. Distribution des fréquences relatives des entités nommées “Date” entre les corpus d’articles associés aux “Robots” et “Algorithmes”



Source : auteurs.

d'anxiété ontologique, comme dans le cas du transhumanisme (Singler, 2019, 2020), donne naissance à des tableaux qui mêlent l'épouvante à l'espérance d'une résurrection de l'humanité.

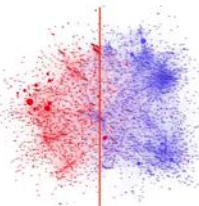
Prophétie et critique

L'opposition entre les deux espaces sémantiques dessine des régimes d'énonciation distincts qui cohabitent au sein de l'espace médiatique. Le premier, prophétique, s'exprime dans le registre du conflit et de la guerre et n'ouvre pas de perspective vers une diplomatie ou une politique permettant d'apaiser la tension entre les humains et les machines. Les articles ne font apparaître les camps en présence que sous la forme d'une asymétrie dans la course aux qualités respectives des humains et des machines. L'alerte anxieuse adressée à la société est sans appel : elle annonce le moment inexorable du dépassement et prophétise le remplacement. Lorsque l'on observe les institutions et ONG les plus mentionnées dans l'espace "Robots", seuls les militaires (*Pentagon, Project Maven, Defense Department, Rand, etc.*) et les organisations de défense de droits de l'homme (*Human Right Watch*) participent à cette dramaturgie des temps futurs (figure 13). Même si certains termes apparaissent en réalité à la frontière entre les deux espaces, le régime d'énonciation de la critique des Algorithmes est à l'inverse éminemment politique et juridique. Les discours appellent constamment une régulation visant à limiter les injustices et à contrôler les nuisances (*disinformation, censure, discrimination, surveillance*). Les problèmes causés par les Algorithmes ne constituent pas une violence univoque, mais un ensemble d'opérations précises et motivées qui appellent une régulation (*government, Senate, White House*) en mobilisant des acteurs spécifiques (*police, High school, FTC, NIST*), des

Figure 13. Extrait des termes les plus cooccurents au sein du graphe relatif aux entités catégorisées "Institutions" dans les deux espaces sémantiques "Robots" et "Algorithmes"

Robots

Pentagon, Congress, Project Maven, Human Rights Watch, Defense Department, Kaist, Allen Institute, NSA, OECD, Rand, WEF, Heritage Foundation, Euro NCAP, Bank of America, IPPR, Defense Innovation Board, Advisory Council, Foreign Office, FCC, human rights organizations, EPA, UN report



Algorithms

government, police, school, Senate, NHS, White House, police departments, ACLU, High School, FTC, Darpa, FBI, South Wales Police, Metropolitan Police, NIST, Big Brother Watch, EPIC, NSPCC, intelligence agencies, Georgetown Law, Princeton, CIA, Washington County, Illinois law, constituency

Source : auteurs.

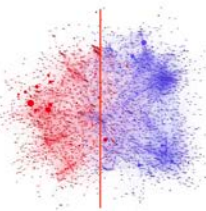
ONG spécialisées (*ACLU, EPIC, Big Brother Watch, NSPCC*) et une expertise académique, juridique notamment (figure 13). La critique des algorithmes fait émerger un ensemble de problèmes publics qui se sont désormais, pour beaucoup d’entre eux, constitués comme des espaces spécialisés rassemblant autour de la mise en évidence d’un type de nuisance acteurs publics, ONG, entreprises et avocats.

Cette tension entre prophétie et critique apparaît aussi dans les listes de personnalités identifiées dans chaque sous-espace – même si certains positionnements sont plus flous, *supra*, note 16 – où s’opposent des futurologues et des chefs d’entreprises visionnaires, d’une part, à des acteurs politiques et économiques et des chercheurs en informatique et en sciences sociales, d’autre part (figure 14). Le débat critique autour de la société algorithmique fait principalement intervenir des acteurs politiques contemporains (*Donald Trump, Hillary et Bill Clinton, Nancy Pelosi, Alex Jones*) et des CEO des grandes entreprises offrant des services numériques largement déployés dans nos sociétés (*Mark Zuckerberg, Sundar Pichai, Jack Dorsey*). Sous le patronage d’Isaac Asimov, le sous-espace Robots est lui dominé par la personnalité d’Elon Musk, incarnation de l’innovateur tourné vers les défis de la conquête spatiale. D’anciens entrepreneurs du numérique (*Dag Kittlaus, Eric Schmidt, Steve Wozniak, etc.*) apparaissent aussi comme des relais de la prophétie, même si une des particularités énonciatives de leurs discours est de dénoncer sur un ton apocalyptique une catastrophe à venir dont ils semblent implicitement souhaiter l’avènement – cette ambivalence est, en France, caractéristique des interventions publiques de Laurent Alexandre (Lassaube, 2019). Essayistes, futurologues et prospectivistes mobilisés par les articles du sous-espace Robots ont pour caractéristiques d’être apparu sur la scène publique lors de la seconde

Figure 14. Extrait des termes les plus cooccurrents au sein du graphe relatif aux entités personnalités catégorisées “Person” dans les deux espaces sémantiques “Robots” et “Algorithmes”

Robots

Elon Musk, Nick Bostrom, Noel Sharkey, Max Tegmark, Eric Schmidt, Dag Kittlaus, Steve Wozniak, Isaac Asimov, Guduruth Banavar, Jürgen Schmidhuber, Toby Walsh, Alexandria Ocasio-Cortez, John Maynard Keynes, Stuart Armstrong, Erik Brynjolfsson, Adrew McAfee, Oren Etzioni, Geoff Hinton, Alexander Reben, Mark Carney, Luciano Floridi, Adam Cheyer, Stuart Russell, Chris Urmson, Gill Pratt, Nigel Shadbolt



Algorithms

Donald Trump, Mark Zuckerberg, Michal Kosinski, Hillary Clinton, Sundar Pichai, Rana El Kaliouby, Guillaume Chaslot, Paul Ekman, Jack Dorsey, Jeffrey Epstein, Alex Jones, Joy Buolamwini, Satya Nadella, Magnus Carlsen, Bill Clinton, Paul Todorov, Sheryl Sandberg, Michelle Obama, Nancy Pelosi, Max Deutsch, Bernie Sanders, Angela Merkel, Jesus, Ziad Obermeyer, Christina Hitrova, Aviv Ovadya, James Barrett

Source : auteurs.

vague “symbolique” de l’IA et du développement du courant transhumaniste (*Nick Bostrom, Noel Sharkey, Toby Walsh, Nigel Shadbolt, Andrew McAfee, Hans Moravec, etc.*). C’est en quelque sorte une génération “ancienne” de la prophétie de l’IA qui, profitant de la troisième vague “connexionniste”, a été remise en selle par les médias.

DANS L’OMBRE DES ROBOTS

La cohabitation des deux régimes discursifs de la prophétie et de la critique dans les débats médiatiques autour de l’IA constitue une manière empirique de montrer comment se structure l’imaginaire des technologies dites « intelligentes ». Cette exploration des discours médiatiques fait apparaître un système de représentation à la fois clivé et interdépendant. Les différences très apparentes entre les mondes des Robots et des Algorithmes témoignent d’abord de la profonde désarticulation entre le futur projeté et les réalités sociotechniques actuelles de l’IA. Alors que l’imaginaire de la disparition de l’humanité nourrit toujours plus activement la science-fiction (de *2001, l’odyssée de l’espace* à *Her*), la discussion des essayistes et certains articles de presse, les chercheurs les plus en vue du domaine, comme Yann Le Cun (2019), ne cessent de souligner que les innovations contemporaines à base de technique d’apprentissage profond et de grands jeux de données n’appartiennent pas à la lignée de technologie qui avaient forgé dans les générations précédentes la rêverie de la machine autonome. Peu attentifs à la variété des lignées technologiques réunies sous le nom d’IA, les médias se révèlent cependant particulièrement prompts à activer un stock de représentations déjà disponibles et riche de nombreuses ressources culturelles et fictionnelles pour relayer la prophétie d’une guerre entre les humains et les robots. On pourrait parler à cet égard d’une sorte d’effet d’*hystérésis* entre l’imaginaire robotique de l’IA et les développements contemporains du *deep learning*. Ce décalage offre un moyen opportuniste de profiter des effets symboliques de la prophétie de vie artificielle pour soutenir la promesse technologique et marchande des algorithmes prédictifs, soutiennent Madeleine Clare Elish et Danah Boyd (2018, p. 58). Il donne au domaine des data science un horizon projectif et une visibilité qui, comme lors des précédentes vagues de l’IA, lui assure l’attention du public et des investisseurs publics et privés.

Si cette interprétation opportuniste est incontestable, il faut cependant souligner les liens d’interdépendances entre ces deux ensembles de représentations de l’IA. Même s’il est parfois très atténué, l’arrière-fond eschatologique d’un

grand conflit entre les humains et les chimères artificielles reste toujours présent dans les débats sur les technologies informatiques dites « intelligentes ». La continuité entre, d'une part, la perte d'autonomie des humains dans le monde algorithmique et, d'autre part, le gain d'indépendance des machines dans celui des robots semble installer un système de vases communicants. Lorsque les algorithmes produisent un ensemble de nuisances qui se superposent l'une sur l'autre (la désinformation, la menace sur la vie privée, le contrôle attentionnel, la censure de l'information, etc.), la critique relève chaque fois que les individus perdent de l'autonomie face aux systèmes socio-techniques qui les ensèrent. Le discours critique mobilisé dans le sous-espace Algorithmes rend d'abord responsables les entreprises du capitalisme numérique de cette nouvelle forme d'exploitation. Mais, il laisse aussi toujours une place à l'idée que celle-ci est renforcée par la difficulté à exercer un contrôle sur des outils qui ont une forme d'existence propre. À l'état de vestige, l'imaginaire du robot autonome refait surface dans la critique des algorithmes sur le mode de l'incontrôlabilité. Les thématiques de la boîte noire, de l'explicabilité des algorithmes (Burrell, 2016), les tropes d'exagération sur le volume, le pouvoir qui sont aujourd'hui associés à tous les discours publics sur les algorithmes témoignent du fait que l'imaginaire d'une machine que personne, pas même leurs concepteurs, ne parvient à domestiquer sont présents dans les deux sous-espaces. La mise en scène d'une logique propre du calculateur sert alors d'arrière-plan à la dramatisation des enjeux et contribue fortement à la sensibilisation des acteurs publics, des médias et du public. On peut faire l'hypothèse que cet imaginaire sous-jacent contribue à contaminer d'une forme de déterminisme technologique les interprétations critiques qui sont conduites pour dénoncer les conséquences du calcul algorithmique sur nos sociétés (Campolo et Crawford, 2020). L'emphase hyperbolique qui nourrit la mobilisation actuelle des ONG, des médias et des pouvoirs publics autour des dangers de l'intelligence artificielle agite toujours, en animant les ressorts d'un imaginaire collectif à disposition, l'idée que des apprentis sorciers sont en train de donner naissance à des vies artificielles dont on ne connaît ni la force ni les intentions.

 RÉFÉRENCES

- ALGORITHM WATCH (2020), « AI Ethics Guidelines Global Inventory », *Algorithm Watch*, April 30, [En ligne] disponible à l'adresse : <https://inventory.algorithmwatch.org/>, consulté le 3 mars 2022.
- BARDINI T. (2000), *Bootstrapping: Douglas Engelbart, coevolution, and the origins of personal computing*, Redwood, Stanford University Press.
- BAREIS J., KATZENBACH C. (2021), « Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and their Performative Politics », *Science, Technology & Human Values*, July.
- BENBOUZID B., MÉNECOEUR Y., SMUHA N. (2022), « Quatre nuances de régulation de l'intelligence artificielle. Une cartographie des conflits de définition », *Réseaux*, n° 232-233, p. 27-64.
- BOLTANSKI L., CLAVERIE E. (2007), « Du monde social en tant que scène d'un procès », in OFFENSTADT N., VAN DAMME S., *Affaires, scandales et grandes causes : de Socrate à Pinochet*, Paris, Stock, p. 395-452.
- BRENNEN, J., HOWARD P., NIELSEN R. (2018), *An industry-led debate: how UK media cover artificial intelligence*, Reuters Institute for the Study of Journalism, December 2018.
- BRYNJOLFSSON E., MCAFEE A. (2014), *The Second machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies*, New York, W. W. Norton & Company.
- BUCHER T. (2016), « The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms », *Information, Communication & Society*, vol. 20, n° 1, p. 30-44.
- BURRELL J. (2016), « How the machine 'thinks': Understanding opacity in machine learning algorithms », *Big Data & Society*, January-June, p. 1-12.
- CAMPOLO A., CRAWFORD K. (2020), « Enchanted Determinism: Power without Responsibility in Artificial Intelligence », *Engaging Science, Technology, and Society*, vol. 6, p. 1-19.
- CARDON D., COINTET J., MAZIÈRES A. (2018), « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle », *Réseaux*, n° 211, p. 173-220.
- CARDON D., CREPEL M. (2019), « Algorithmes et régulation des territoires », in LE GALÈS P., COURMONT A. *Gouverner la ville numérique*, Paris, PUF, p. 81-99.
- CHATEAURAYNAUD F. (2013), « Regard analytique sur l'activité visionnaire », in BOURG D. (dir.), *Du risque à la menace. Penser la catastrophe*, Paris, PUF, p. 287-310.

- CHATEAURAYNAUD F. (2014), « Trajectoires argumentatives et constellations discursives. Exploration socio-informatique des futurs vus depuis le nanomonde », *Réseaux*, n° 188, p. 121-158.
- CHATEAURAYNAUD F., DEBAZ J. (2019), « Agir avant et après la fin du monde, dans l'infinité des milieux en interaction », *Multitudes*, vol. 3, n° 3, p. 126-132.
- CREPEL M., DO S., COINTET J., CARDON D., BOUACHERA Y. (2021), « Mapping AI issues in media through NLP methods », *Proceedings of the Conference on Computational Humanities Research (CHR2021)*, p. 77-91, [En ligne] disponible à l'adresse : http://ceur-ws.org/Vol-2989/long_paper22.pdf, consulté le 3 mars 2022.
- CREVIER D. (1997), *À la recherche de l'intelligence artificielle*, Paris, Champs/Flammarion [1^{re} éd. américaine 1993].
- ELISH M.C., BOYD D. (2018), « Situating methods in the magic of Big Data and AI », *Communication Monographs*, vol. 85, n° 1, p. 57-80.
- FJELD J., NELE A., HANNAH H., ADAM N., MADHULIKAS. (2020), « Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center for Internet & Society ».
- FLORIDI L. (2019), « Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical », *Philosophy & Technology*, p. 185-193.
- FLICHY P. (2001), *L'imaginaire d'internet*, Paris, La Découverte.
- GANASCIA J.G. (2017), *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle ?*, Paris, Le Seuil.
- GRUDIN J. (2009), « AI and HCI: Two fields divided by a common focus », *AI Magazine*, vol. 30, n° 4, p. 48-57.
- HARARI Y.N. (2017), *Homo Deus : une brève histoire de l'avenir*, Paris, Albin Michel.
- HAUGELAND J. (1985), *Artificial Intelligence: The Very Idea*, Cambridge, Massachusetts: MIT Press.
- HAYLES K. (1999), *How We Became Posthuman. Virtual Bodies, Cybernetics, Literature and Informatics*, Chicago, The University of Chicago Press.
- JASANOFF S. (2015), « Future Imperfect: Science, Technology, and the Imaginations of Modernity », in JASANOFF S., KIM S.-H., *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Chicago, University of Chicago Press.
- JOBIN A., IENCA M., VAYENA E. (2019), « The global landscape of AI ethics guidelines », *Nature Machine Intelligence*, vol. 1, n° 9, p. 389-399.
- KATZ Y. (2017), « Manufacturing an artificial intelligence revolution », [En ligne] disponible à l'adresse : SSRN 3078224.

- LASSAUBE G. (2019), « Cassandre appelant de ses vœux la catastrophe : quand Laurent Alexandre écrivait de la science-fiction », *Zisel*, n° 6, octobre, p. 389-416.
- LE CUN Y. (2019), *Quand la machine apprend. La révolution des neurones artificiels et de l'apprentissage profond*, Paris, Odile Jacob.
- LEE K. (2018), *AI Super-Powers. China, Silicon Valley and the World Order*, Boston, Houghton Mifflin Harcourt.
- MARCUS G., DAVIES E. (2019), *Rebooting AI. Building Artificial Intelligence We Can Trust*, New York, Pantheon Books.
- MARKOFF J. (2015), *Machines of loving grace. Between human and robots*, New York, HarperCollins Publishers.
- MARTIN C. D. (1993), « The Myth of the Awesome Thinking Machine », *Communications of the ACM*, vol. 36, n° 4, p. 12-133.
- MCCORDUCK P. (2004), *Machine who think*, Natick, A. K. Peters.
- METZ C., (2021), *Genius Makers. The Maverick who brought to Google, Facebook and the World*, New York, Penguin Random House LLC.
- MEUNIER A., RICCI D., CARDON D., CREPEL M. (2019), « Les glitches, ces moments où les algorithmes tremblent », *Techniques & Culture*, n° 72, p. 200-203.
- MITTELSTADT B. (2019), « Principles Alone Cannot Guarantee Ethical AI », *Nature Machine Intelligence*, November, [En ligne] disponible à l'adresse : <https://ssrn.com/abstract=3391293>, consulté le 3 mars 2022.
- MITTELSTADT B., ALLO P., TADDEO M., WACHTER S., FLORIDI L. (2016), « The ethics of algorithms: Mapping the debate », *Big Data & Society*, vol. 3, n° 2, p. 1-21.
- MOSCO V. (2005), *The Digital Sublime: Myth, Power and Cyberspace*, Cambridge, The MIT Press.
- NATALE S., BALLATORE A. (2020), « Imagining the Thinking Machine: Technological Myths and the Rise of Artificial Intelligence », *Convergence: The International Journal of Research into New Media Technologies*, vol. 26, n° 1, p. 3-18.
- OCDE (2019), *L'intelligence artificielle dans la société*, Paris, Éditions OCDE.
- PERRAULT R., SHOHAMY., BRYNJOLFSSON E., CLARK J., ETCHEMENDY J., GROSZ B., LYONS T., MANYIKA J., MISHRA S., NIEBLES J.C. (2019), « The AI Index 2019 Annual Report », *AI Index Steering Committee, Human-Centered AI Institute*, Stanford University, Stanford, CA, December.
- RECCHIA G. (2020), « The Fall and Rise of AI: Investigating AI Narratives with Computational Methods », in CAVE S., DIHAL K., DILLON S., *AI, Narratives: A History of Imaginative Thinking about Intelligent Machines*, Oxford, Oxford University Press.

RUSSELL S. J., NORVIG P., CANNY J. F. (2010), *Artificial Intelligence: A Modern Approach*, Saddle River, Pearson Education.

SEAVER N. (2017), « Algorithms as culture: Some tactics for the ethnography of algorithmic systems », *Big Data & Society*, July-December, p. 1-12.

SINGLER B. (2019), « Existential hope and existential despair in AI apocalypticism and transhumanism », *Zygon*, vol. 54, n° 1, p. 156-176.

SINGLER B. (2020), « Blessed by the algorithm: Theistic conceptions of artificial intelligence in online discourse », *AI & Society*, n° 35, p. 945-955.

TEGMARK M. (2017), *Life 3.0: being human in the age of artificial intelligence*, New York, Knopf.

TENNANT C., STILLGOE J. (2021), « The Attachments of ‘autonomous’ vehicles », *Social Studies of Science*, August 14.

VINSEL L., RUSSELL A.L. (2020), *The Innovation Delusio. How Our Obsession with the New Has Disrupted Work That Matters Most*, New York, Currency.

ANNEXE 1

Liste des sources extraites sur Factiva

Presse Nationale USA

ABCNEWS : <http://abcnews.go.com/>

Barron's : <https://www.barrons.com>

Business Insider : <https://www.businessinsider.com>

CBS News : <http://www.cbsnews.com/>

Chicago Tribune : <http://www.chicagotribune.com/>

CNN : <http://www.cnn.com/>

Dailynews : <https://www.nydailynews.com/>

Forbes : <https://www.forbes.com>

Fox News : <http://foxnews.com/>

Los Angeles Times : <http://www.latimes.com/>

MSNBC News : <http://www.msnbc.msn.com/>

New York Newsday : <http://www.newsday.com/>

New York Post : <http://www.nypost.com/>

New York Times : <http://www.nytimes.com/>

Newsweek : <https://www.newsweek.com>

Politico : <https://www.politico.com>
 Rochester Democrat Chronicle : <http://www.democratandchronicle.com/>
 Slate : <https://slate.com>
 The Huffington Post : <https://www.huffingtonpost.com>
 The Verge : <https://www.theverge.com>
 Time Magazine : <http://www.time.com>
 USA Today : <https://www.usatoday.com/>
 Vice News : <https://www.vice.com/>
 Wall Street Journal Online : <http://online.wsj.com/>
 Wall Street Journal : nc
 Washington Post : <http://www.washingtonpost.com/>
 Washington Times : <http://www.washingtontimes.com/>

Presse Nationale UK

Daily Mail : <https://www.dailymail.co.uk/>
 Daily Record : <http://www.dailyrecord.co.uk/>
 Daily star : <https://www.dailystar.co.uk>
 Economist : <https://www.economist.com>
 Evening Standard : <https://www.standard.co.uk>
 Evening Times : <http://www.eveningtimes.co.uk/>
 Express : <https://www.express.co.uk>
 Guardian : <http://www.guardian.co.uk/>
 Independent : <http://www.independent.co.uk/>
 Mail online : nc
 Morning Star : <https://morningstaronline.co.uk>
 Observer : <http://www.observer.co.uk/>
 Sunday Mail : <http://www.sundaymail.co.uk/>
 Telegraph : <http://www.telegraph.co.uk/>
 The Mirror : <http://www.mirror.co.uk/>
 The Sunday Mirror : nc
 The Sun : <http://www.the-sun.co.uk/>
 The Sunday Times : <http://www.sunday-times.co.uk/>
 The Times : <http://www.thetimes.co.uk/>
 Which : <https://www.which.co.uk>