L'Intelligence Artificielle, une approche intersectionnelle

Réflexions sur l'éthique et la justice sociale de l'IA

Artificial Intelligence, an intersectional approach: Thinking about the ethics and social justice of AI

< Julie MARQUES [1] >

1. Laboratoire Prefics, Université Rennes 2
juliemarques@hotmail.fr

DOI : 10.25965/interfaces-numeriques.4796

< RÉSUMÉ >

Dans cet article, nous proposons d'analyser l'IA et son éthique au prisme d'une approche intersectionnelle, afin de dépasser l'idée que cette Technique serait neutre, pour la penser comme éminemment sociale, idéologique et politique. À partir d'un corpus de soft law sur la gouvernance éthique de l'IA, composé de discours institutionnels émis par des instances aux enjeux hétérogènes, nous interrogeons les principes, concepts qui cirsculent dans ces discours, ainsi que les limites de l'éthique appliquée à la Technique dite intelligente. En parallèle, nous ouvrons un espace de réflexion sur les enjeux de justice sociale imbriqués dans l'IA et son éthique.

< MOTS-CLÉS >

Intelligence Artificielle, intersectionnalité, principes éthiques, justice sociale

< ABSTRACT >

In this article, we propose to analyze AI and its ethics through the prism of an intersectional approach, in order to go beyond the idea that this Technique would be neutral, to think of it as eminently social, ideological and political. From a corpus of soft law on the ethical governance of AI, composed of institutional speeches issued by bodies with heterogeneous stakes, we question the principles and concepts that circulate in these speeches, as well as the limits of ethics applied to the so-called intelligent Technique. In parallel, we open a space of reflection on the stakes of social justice embedded in AI and its ethics.

< KEYWORDS >
Artificial Intelligence, intersectionality, ethical principles, social justice

Derrière un ensemble de discours sur l'Intelligence Artificielle (IA) se cachent horizons des possibles, utopies, catastrophes et autres apocalypses, formant des techno-imaginaires, c'est-à-dire du « *fonctionnel et du fictionnel* », de la Technique et du discours (Musso, 2020, 116). Derrière la Technique, l'idée de transformation ou de progrès, souvent corrélée à un mieux vivre, une évolution forcément positive pour l'humanité et à la place de ces espoirs une réalité parfois tout autre. Sans entrer pour autant dans l'écueil d'une technophobie, force est de constater que la Technique dite intelligente présente le risque de reproduire et de systématiser les formes d'oppression (Noble, 2018), sans une réelle réflexion sur la gouvernance des algorithmes au prisme d'une approche intersectionnelle (Bilge, 2009).

Dans cet article, nous proposons d'analyser l'IA et son éthique au prisme d'une approche intersectionnelle, afin de dépasser l'idée que cette Technique serait neutre, pour la penser comme éminemment sociale, idéologique et politique. À partir d'un corpus de *soft law* sur la gouvernance éthique de l'IA, composé de discours institutionnels émis par des instances aux enjeux hétérogènes, nous interrogeons les principes, concepts qui circulent dans ces discours, ainsi que les limites de l'éthique appliquée à la Technique dite intelligente. En parallèle, nous ouvrons un espace de réflexion sur les enjeux de justice sociale imbriqués dans l'IA et son éthique.

1. Méthodologie et corpus

La recherche s'ancre dans un cadre théorique issu des sciences du langage, des sciences de l'information communication et s'appuie particulièrement sur les études de genre.

Le discours, qui est le matériel de notre analyse, est pensé comme forme de savoir/pouvoir. Il est perçu comme essentiel dans la construction des idéologies, des savoirs, du monde social et des rapports qui s'y inscrivent (Bourdieu, 2001 ; Foucault, 1969). La Technique est construite par cet ensemble de discours circulants issus de la société

civile, des gouvernements ou encore des entreprises. Pour effectuer l'analyse des discours (AD), un corpus de 115 recommandations, en anglais, en matière de gouvernance éthique de l'IA a été constitué. Le corpus comprend des publications qui vont de 2011, date d'apparition de ce type de document à mars 2021, date à laquelle la collecte a pris fin. Les publications peuvent être qualifiées de discours institutionnels en matière d'IA, mais sont souvent définies comme des « lignes directrices » ou « recommandations » pour une IA éthique. Ces recommandations forment un *interdiscours* en mutation constante, c'est-à-dire un vaste ensemble de discours qui ne sont pas hermétiques les uns aux autres, mais s'influencent, se répondent et participent « d'une chaine verbale interminable. » (Maingueneau, 2014, 23). Ils constituent un corpus de *soft law* dont se dotent diverses instances et supposent une hétérogénéité de formes et d'effets, mais sont caractérisés par une diminution de la pression législative (Lavergne, 2013). Pour garantir une cohérence du corpus, les recommandations sur une branche spécifique de l'IA n'ont pas été retenues. L'ensemble des publications ont été triées par type d'organisme producteur, date de publication et pays de production. À partir de cette répartition, une cartographie des discours pour la gouvernance éthique de l'IA a été faite, en reprenant la méthodologie proposée par A. Jobin et al (2019). Les types d'organismes producteurs sont répartis comme suit :

- Institut de recherche (9 %)
- Fondations (3 %)
- Multipartites (8 %)
- Instances d'autorité publique, comprennent toutes les publications gouvernementales, intergouvernementales, supranationales ou relevant du secteur public (42 %)
- Association,
- Privé, comprend les organisations du secteur privé (13 %)
- N.A, pour « non available », qui correspond aux organisations dont le statut est inconnu (3 %)

Notre corpus de texte a été encodé en conservant cette répartition, pour être analysé via le logiciel textométrique IRAMUTEQ. La

textométrie : « permet une observation à la fois fine et globale des textes, tout en restant proche de la matérialité de ces derniers par le recours à des instruments d'accès aux données, et en mettant en valeur la réalité langagière […] » (Saigh et al., 2017, 164). Diverses fonctionnalités sont utilisées pour rendre compte du corpus statistiquement et visuellement. Notamment, l'index hiérarchique, qui donne la fréquence des occurrences, l'Analyse Factorielle des Correspondances (AFC), qui est une méthode de statistiques descriptives pour l'analyse de données textuelles. Elle « identifie les faits saillants d'un corpus en termes de distribution du stock lexical. » (Née, 2017, 152). À partir de l'AFC, on constate les spécificités lexicales pour chaque catégorie encodée. La méthode Reinert décrit les lois de distribution des occurrences, du lexique d'un corpus. Elle met en évidence les mondes lexicaux du corpus, en calculant des classes. Cette fonctionnalité met en évidence des segments représentatifs de la classe lexicale, calculés « en fonction de l'association statistique forte d'un segment à une classe lexicale » (Saigh et al., 2017, 168). Enfin, nous avons recours à l'analyse des similitudes, qui est une représentation graphique des cooccurrences. Elle « présente graphiquement la structure d'un corpus en distinguant les parties communes et les spécificités des variables codées, ce qui permet de mettre en avant la relation entre les différentes formes dans les segments de texte. » (Saigh et al., 2017, 168). Cet outil et ses fonctionnalités rendent possible une description détaillée et systématique de grand corpus de textes.

2. Intelligence Artificielle et les rapports sociaux de sexe, race[1] et classe : une co-construction

Dans ce chapitre, nous définissons l'IA, tout en montrant que sa formalisation est imprégnée de divers techno-imaginaires. Nous théorisons ensuite la manière dont l'IA et les rapports sociaux de sexe, de

1 Le terme « race » est utilisé dans un discours critique, il n'est absolument pas évoqué dans un sens naturaliste ou biologiste qui supposerait la hiérarchie entre des groupes humains, notamment dans la manifestation de différences physiques, cette hiérarchie n'existe pas. Il est utilisé pour venir nommer : « *un rapport de pouvoir qui structure, selon des modalités diverses en fonction des contextes et des époques, la place sociale assignée à telle ou telle groupe au nom de ce qui est censé être la radicale altérité de son origine.* » (Mazouz, 2020, 26).

race et de classe se co-construisent, tout en soulignant la nécessité d'une approche intersectionnelle pour comprendre comment l'IA perpétue des inégalités, oppressions et marginalisations.

2.1. L'Intelligence Artificielle : Définition et techno-imaginaires

L'expression « Intelligence Artificielle » cristallise déjà les premières inquiétudes et imaginaires ; qu'ils soient de type prophétique ou apocalyptique. Elle suppose une forme d'intelligence, calquée sur celle perçue comme « suprême » : l'intelligence humaine (Doueihi, 2019). Si l'appellation est souvent remise en cause par les expert.e.s, qui trouvent qu'il s'agit d'une exagération des capacités que l'on peut accorder au computationnel, cela n'empêche pas pour autant les comparaisons avec les modes de réflexion humains. Cette comparaison s'accentue avec le développement des nouvelles méthodes de modélisation, telle que le *machine learning*, qui suppose que la machine apprend, ou encore le *deep learning*, souvent comparé au réseau neuronal du cerveau humain (Doueihi, 2019).

Techniquement, l'IA est : « un corpus de concepts et de techniques permettant à une machine de réaliser des tâches au moyen de programmes informatiques, simulant parfois ainsi, dans une certaine mesure, l'intelligence humaine. » (Bertail et al, 2019, 6). L'IA peut dépendre d'algorithmes, c'est-à-dire de modélisation du réel, de solutions mathématiques qui prennent la forme : « [d'] un ensemble de règles et d'instructions écrites en vue d'obtenir un résultat. » (Bertail et al, 2019, 6).

L'IA est ancrée dans un ensemble d'imaginaires cybernéticiens, qui a influencé sa conceptualisation, puis sa formalisation et elle est affiliée à l'idée que l'humain devenu aussi puissant que Dieu créerait un être à son image (Musso, 2020). Idée qui transparaît dans l'ouvrage de N. Wiener sur la cybernétique, dans lequel, reprenant la célèbre expression biblique, il écrit : « *L'homme fait l'homme à son image.* » (Wiener, 1964 cité par Collet, 2019, 67). Cet imaginaire d'un auto-engendrement transparaît dans de multiples conceptualisations de l'IA de N.Wiener à J. Von Neumann en passant par A. Turing (Collet, 2019).

Cependant, si l'on regarde de plus près les recherches et matérialisations de l'IA, on note les limites de cette « intelligence ». Ainsi l'idée de créer un « être » conscient et autonome n'a pas abouti et l'IA prend plutôt la forme d'une prothèse ou un moyen « d'augmenter » l'humain, de l'accompagner dans ces pratiques et décisions (Cardon, 2019). En cela, l'IA et les algorithmes appartiennent à la Technique, comme décrit par J. Ellul, c'est-à-dire comme phénomène généralisé où l'efficacité et la performance prévalent sur toutes autres valeurs dont les libertés humaines (Ellul, 1954).

Ce court détour par la genèse de l'IA, permet de comprendre les techno-imaginaires cybernéticiens qui imprègnent la formalisation de cette Technique. Nous poursuivons cette réflexion en analysant l'IA au prisme d'une approche intersectionnelle, pour dépasser l'idée reçue que l'IA serait neutre.

2.2. L'Intelligence Artificielle à l'intersection des rapports sociaux

Parce qu'il y a une insinuation constante que la Technique serait neutre notamment de par son caractère mathématique, on s'imagine bien souvent qu'elle serait objective et comme en dichotomie avec le monde social (Akrich, 1994 ; Haraway, 2007). Pourtant, la Technique est intrinsèquement sociale (Jouët, 2003) et par affiliation l'IA doit être comprise et analysée dans la relation qu'elle entretient avec le monde social.

Tout d'abord, la notion de genre est un des préalables théoriques, au prisme duquel nous appréhendons la Technique. Il fait référence à la construction sociale du sexe et du genre, des rôles et valeurs qu'on leur attribue et qui ne relèvent pas de l'inné (Dorlin, 2008). Il est un processus de différenciation et de hiérarchisation, qui peut être assimilé à un rapport social non-cumulatif et co-produit (Kergoat, 2012). Il est un rapport de pouvoir à l'intersection de diverses formes de marginalisation et systèmes d'oppressions qui varient selon les contextes sociaux, géographiques et historiques (Lépinard et Mazouz, 2021). Le genre est avec les rapports sociaux : « multiples et déterminées simultanément et de façon interactive par plusieurs axes d'organisation sociale significatifs » (Bilge, 2009, 71).

Cette appréhension du genre, articulé aux rapports de pouvoir, s'ancre dans une approche intersectionnelle qui vise à saisir les logiques de domination, en montrant que l'imbrication de plusieurs catégories telles que le sexe, la race, la classe, l'orientation sexuelle, l'âge, le handicap, etc. renvoie à différentes expériences d'oppression et de marginalisation des personnes. L'intersectionnalité permet de sortir de l'appréhension cumulative ou additionnelle des oppressions, qui perpétue l'invisibilisation de certains vécus. Elle a été développée par les féministes noires états-uniennes et formalisée par K. Crenshaw en 1989, afin de montrer que la prise en compte des discriminations au niveau du droit états-unien, de manière séquentielle, invisibilisait les expériences des personnes minorisées. Cette approche a été appliquée à d'autres contextes afin de quitter « un système qui pense la discrimination sexuelle à partir de l'expérience normative des femmes blanches et la discrimination raciale à partir de l'expérience normative des hommes noirs. » (Bilge et Roy, 2010, 57), pour analyser la complexité des expériences micro et macrosociologiques. Avec l'intersectionnalité, on prend « en considération des interactions entre ces catégories multiples qu'elle considère comme mutuellement constitutives » (Bilge et Roy, 2010, 57).

Dans ce travail, la Technique n'est pas pensée comme neutre et fait l'objet d'une critique féministe. Elle est le produit d'une histoire et d'idéologies ; elle est politique, culturelle et sociale (Cockburn, 1981 ; Winner, 1980) et interagit avec le monde social et les groupes sociaux qui le composent. De fait, il y a une co-construction entre Technique et monde social (Haraway, 2020). Cette approche de l'IA dépasse les écueils d'un déterminisme technique, tout autant qu'une dichotomie entre Technique et société, de fait : « […] les techniques sont à la fois "ce qui façonne la société" et "ce que la société façonne". » (Akrich, 1994, 123). Dans cette perspective, l'IA est un miroir de nos sociétés et se trouve à l'intersection des rapports sociaux, qu'elle transforme, reproduit, voire systématise. Les débordements de l'IA ne manquent, d'ailleurs, pas en matière « d'oppressions algorithmiques »[2], notamment parce que nous

2 Traduit de l'anglais : « Algorithmic Oppression ». Nous empruntons l'expression à S. U. Noble (2018) et préférons parler d'oppressions algorithmiques plutôt que de « biais », afin de rappeler que les rapports sociaux, la domination et la marginalisation sont systémiques, ils sont le produit

produisons de la Technique sans interroger qui elle représente ou comment elle impacte les personnes et plus spécifiquement les personnes minorisées, dont elle peut tout autant accentuer la visibilité, qu'invisibiliser et/ou marginaliser. Divers exemples, d'IA reproduisant les rapports sociaux de sexe, race et classe existent. Ainsi des IA réactualisent le racisme dans le cadre judiciaire en calculant les risques de récidives comme plus élevés pour les personnes racisées (Larson et al, 2020), tandis que certains algorithmes perpétuent les discriminations à l'embauche, en excluant les profils des femmes (Hamilton, 2018) et la reconnaissance faciale ignore les femmes racisées[3]. Tandis que les données à partir desquelles ces systèmes sont créés continuent de marginaliser les personnes minorisées, leurs parcours ou vécus (Criado-Perez, 2019 ; Noble, 2018). Penser l'IA au prisme d'une approche intersectionnelle suppose d'interroger les idéologies imbriquées dans cette Technique, la manière dont elle impacte les personnes, systématise les inégalités et oppressions, tout en ouvrant une réflexion sur la justice sociale de/par cette Technique.

Face à ces risques, nombre d'instances en appellent à une gouvernance éthique de l'IA. Les recommandations sur le sujet ouvrent des négociations entre des instances aux enjeux hétérogènes sur ce qui semble n'être qu'une norme technique, mais en réalité renferme du politique et des modes de vie future. De fait, quels principes éthiques ces discours font-ils circuler, aujourd'hui ? Pour quelle justice sociale, demain ?

incessant d'institutions et se jouent à des niveaux aussi bien micro que macrosociologique. Ils supposent des solutions politiques qui engagent la responsabilité de tout.te.s et ne peuvent être ramener à des solutions technologiques ou numériques à des niveaux seulement individuels (Hampton, 2021).

3 « Gender Shades ». Consulté le 11 juin 2020. http://gendershades.org/overview.html.

3. Analyse des discours sur l'éthique de l'Intelligence Artificielle pour penser l'intersectionnalité et la justice sociale

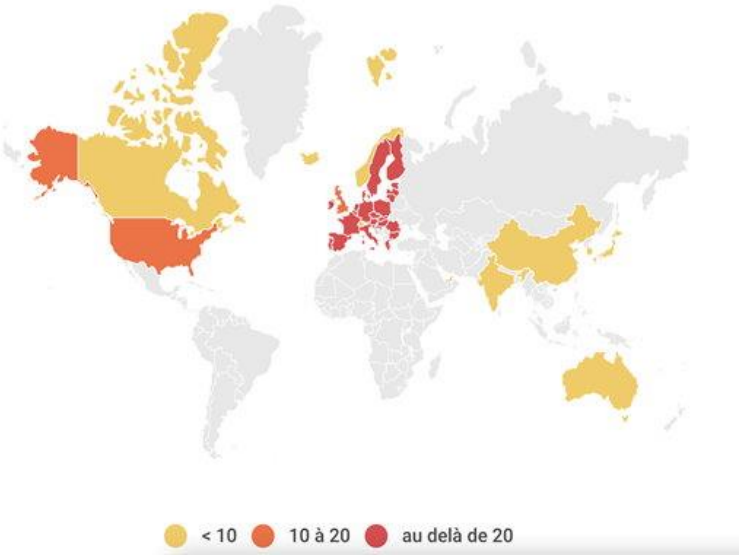
Dans un premier temps, nous décrivons les résultats obtenus via IRAMUTEQ, que nous analysons et mettons en discussion dans un second temps.

3.1. Cartographie et description du corpus

Dans ce chapitre, nous proposons une cartographie et description des résultats obtenus grâce au logiciel IRAMUTEQ.

La cartographie des publications donne une première image macro et quantitative du corpus. Ainsi l'éthique de l'IA est traitée de manière croissante à partir de 2017, avec 78,8 % des publications produites après cette date, bien que le sujet apparaisse dès 2011. Les plus grands producteurs de ces recommandations sont les instances d'autorité publique (41,7 %), ainsi que les organisations du secteur privé (22 %). La production des concepts de l'éthique de l'IA est géographiquement concentrée dans les pays du Nord.

Figure 1. Répartition géographique des publications

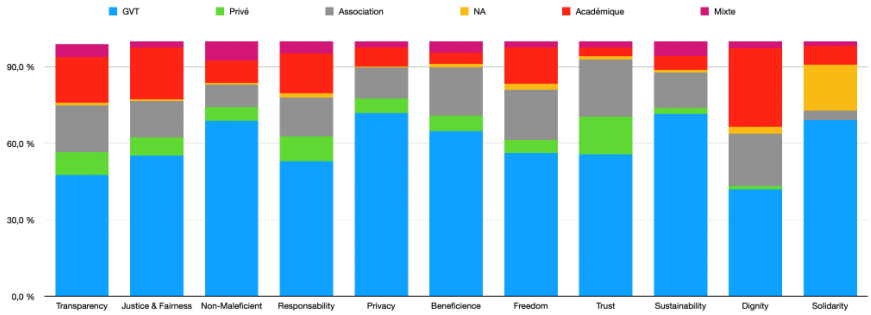


L'analyse textométrique révèle, d'abord, quelques évidences. Ainsi l'index hiérarchique signale la prépondérance des occurrences techniques, comme « AI », « datum », « system ». Toutefois, on note que « human » est la cinquième occurrence. Le nuage de mots suivant met en évidence les termes clés du corpus.

Figure 2. Nuage de mots du corpus

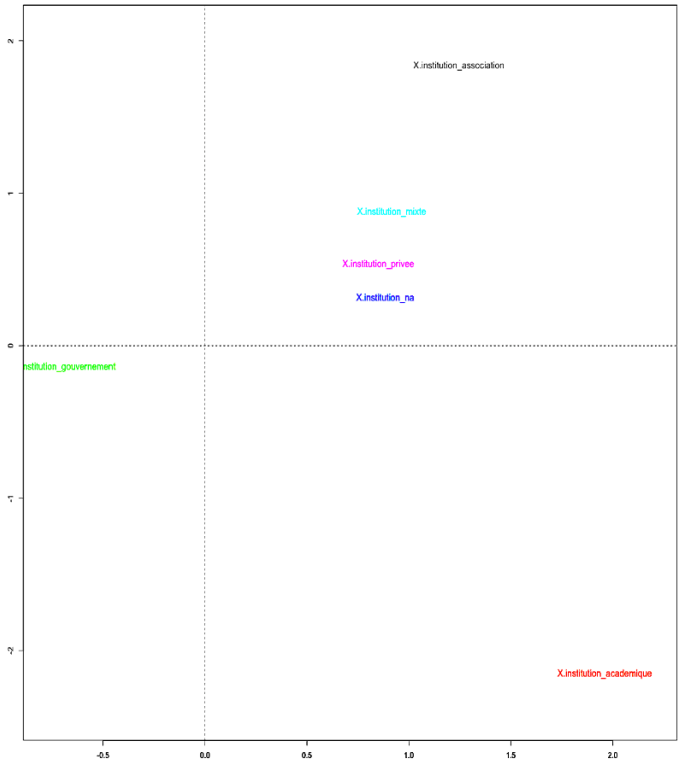
En partant de la typologie des principes de l'IA éthique proposée par A.Jobin et al, nous nous sommes appuyés sur l'outil statistique pour rendre compte de la représentation statistique de ces principes. Sur les onze principes d'IA éthique, celui de « Justice & Fairness » détient l'occurrence la plus importante dans le corpus, suivi par celui de « Non-Maleficient », puis « Responsability ». Tandis que les principes de « Solidarity » et « Dignity » sont ceux avec le moins d'occurrences. La répartition des occurrences dans le corpus s'effectue comme suit :

Figure 3. Répartition des instances qui s'expriment sur les principes éthiques



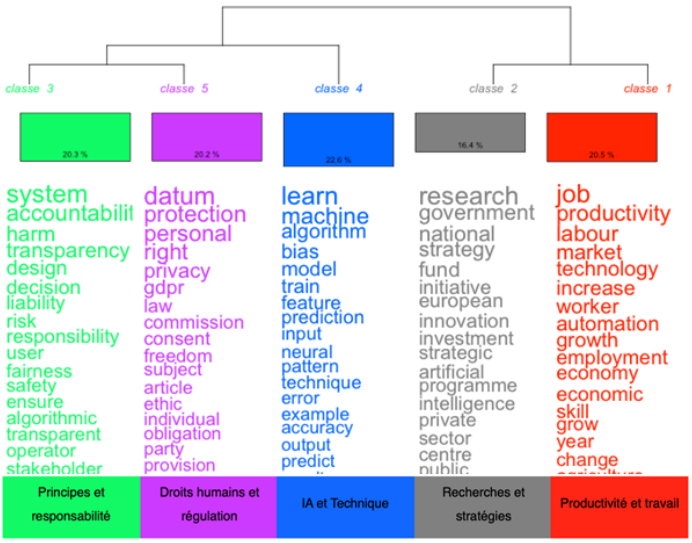
L'AFC révèle les disparités et regroupements entre les catégories d'instances qui s'expriment, tout en rendant compte des spécificités lexicales des catégories d'instance.

Figure 4. AFC du corpus



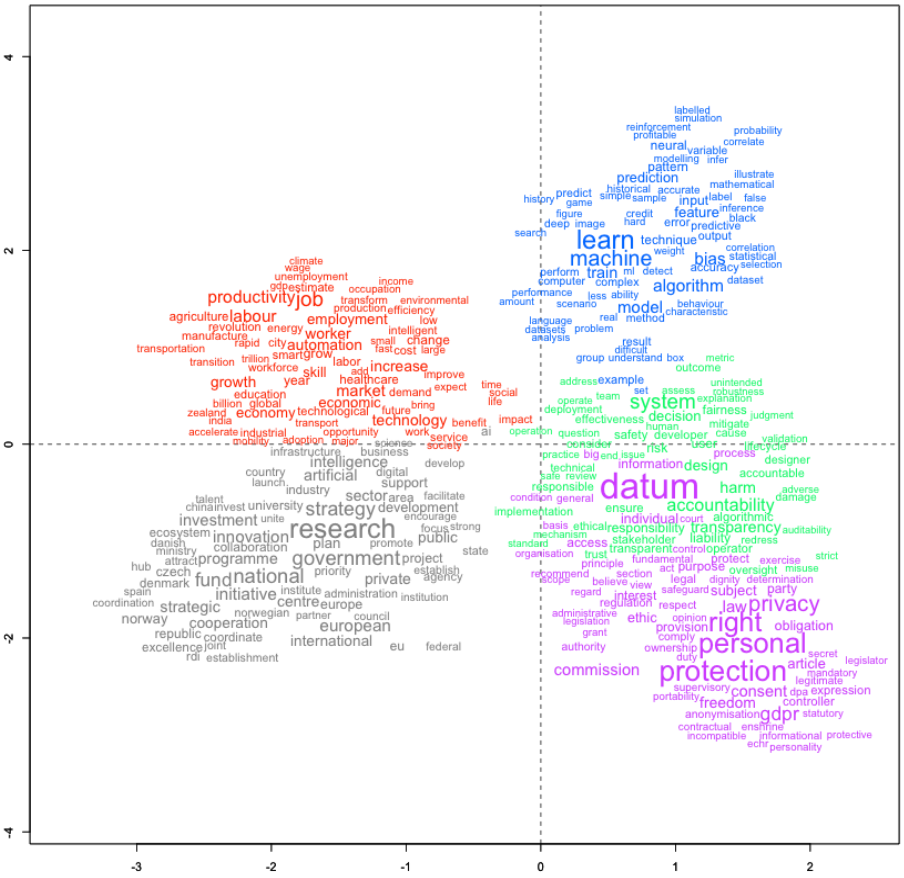
La méthode Reinert montre que le corpus est réparti en cinq classes, qui s'organisent comme suit :

Figure 5. Dendogramme des classes du corpus



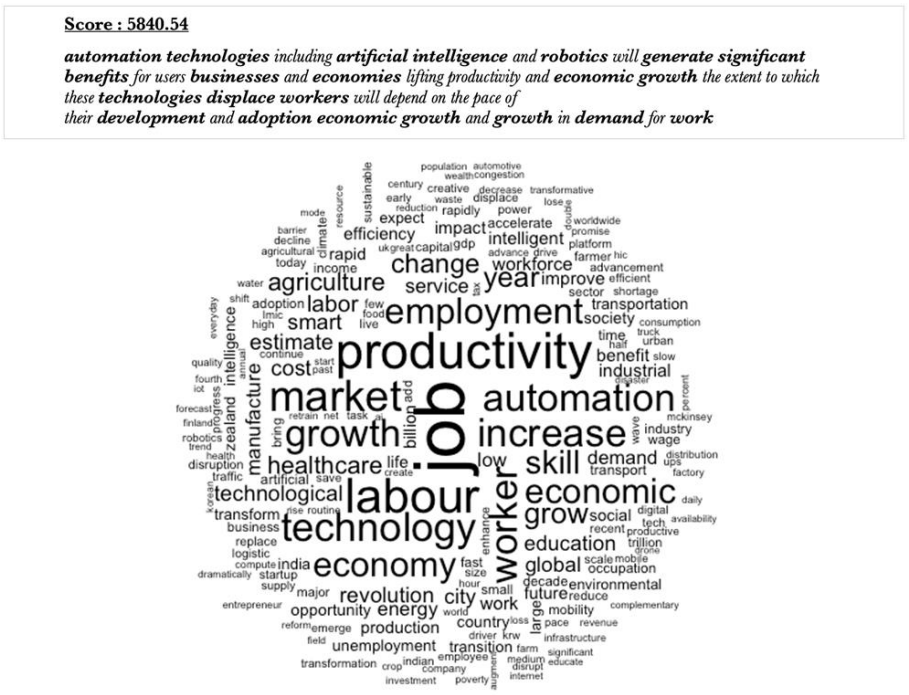
L'AFC des classes montre une proximité des classes 1 et 2, et une autre entre les classes 3, 4 et 5, avec une association entre « Productivité et travail » et « Recherches et stratégies » sur l'axe 2, tandis qu'elles s'opposent à celles de « Principes et responsabilité », « IA et Technique » et « Droits humains et régulation ». Les classes 3 et 5 sont particulièrement associées, ce qui montre un rapprochement entre principes éthiques de l'IA et droits humains.

Figure 6. AFC des classes issues de la méthode Reinert



La classe 1 renvoie à l'idée de potentialité de l'IA sur l'emploi et l'économie que ce soit de manière positive ou négative :

Figure 7. Segment représentatif et nuage de mots de la classe 1



La classe 2 renvoie aux stratégies développées par les instances d'autorité publique pour être acteur dans le développement et le déploiement de l'IA, le segment représentatif témoigne d'une volonté de s'inscrire pleinement dans les enjeux liés à l'IA.

Figure 8. Segment représentatif et nuage de mots de la classe 2

Score : 10902.19

Encouraging member states *to set up* ***national ai strategies outlining investment levels*** *and* ***implementation*** *measures maximizing* ***investments*** *through* ***partnerships fostering investment*** *in* ***strategic ai research*** *and* ***inn ovation*** *through* ***ai public private partnerships*** *and a* ***leaders*** *group as well as a specific* ***fund***

La classe 3 met en évidence les principes qui devraient être appliqués à l'IA, du type d'IA éthique que les instances souhaitent développer.

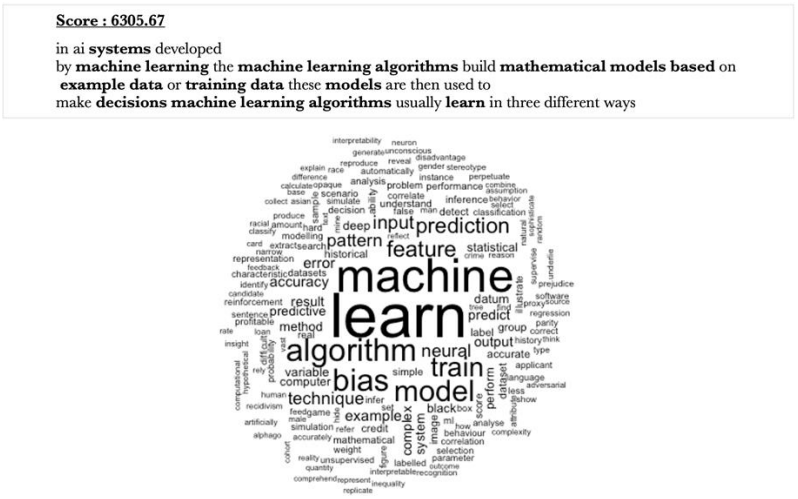
Figure 9. Segment représentatif et nuage de mots de la classe 3

Score : 4871.29

Mechanisms *should be* ***put*** *in* ***place*** *to* ***ensure responsibility*** *and* ***accountability*** *for* ***ai systems*** *and their* ***outcomes*** *this* ***includes*** *both before and after their* ***design*** *development* ***deployment*** *and* ***operation*** *the* ***organisation*** *and* ***individual accountable*** *for the* ***decision*** *should be identifiable as necessary*

La classe 4 renvoie essentiellement aux éléments du discours qui décrivent l'IA, son fonctionnement, ce sont des éléments de définition.

Figure 10. Segment représentatif et nuage de mots de la classe 4

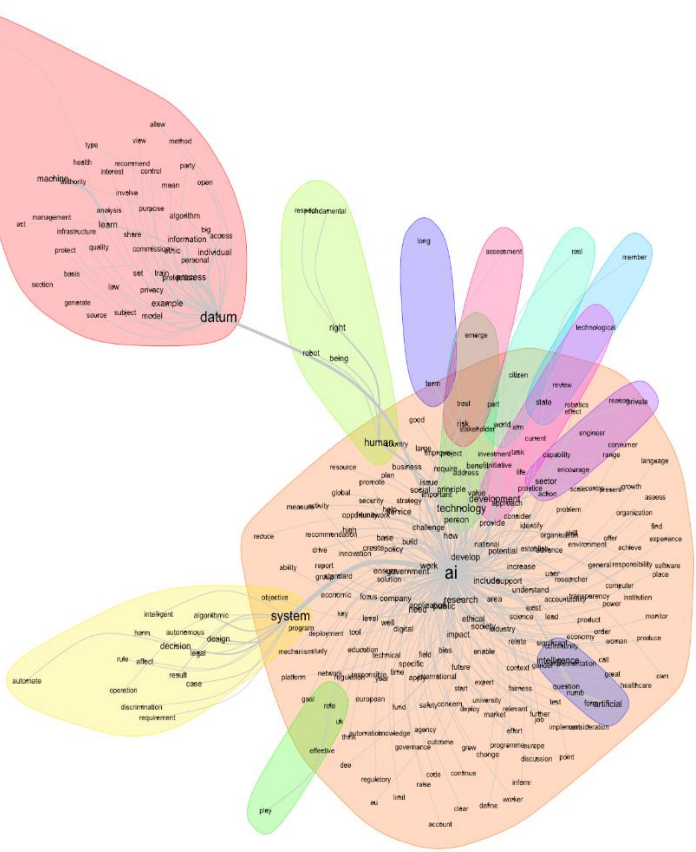


Enfin, la classe 5 qualifie les lois ou les droits fondamentaux existants à respecter lors du développement ou du déploiement d'une IA :

Figure 11. Segment représentatif et nuage de mots de la classe 5

Enfin, le graphe de l'analyse des similitudes montre une homogénéité des discours, avec une forme centrale autour de laquelle le reste des termes gravitent. Le halo principal contient au centre le terme « AI », autour on voit un nuage de mots assez dense et épars avec des terminologies variées, allant de « agency », en passant par « governance » ou « business ». On trouve quelques halos de petite taille directement liés à « AI ». Les deux plus grands halos étant « datum » et « system », vocabulaire qui reste technique.

Figure 12. Analyse des similitudes du corpus



Cette cartographie et description du corpus via IRAMUTEQ met en évidence les données saillantes du corpus. Nous proposons une analyse critique de ces résultats.

3.2. L'éthique et le principe de « fairness » dans l'IA, quelles perspectives ?

La textométrie met en lumière les phénomènes de circulation des principes de l'IA éthique, de même que certains enjeux en termes de contenu, que nous mettons en discussion dans une continuité des réflexions des approches féministes des techno-sciences (Van Zoonen, 1992) et au prisme de l'intersectionnalité, pour questionner qui produit les savoirs/pouvoirs de l'IA et de son éthique (Chabaud-Rychter et Gardey, 2002).

Les recommandations pour une gouvernance éthique de l'IA constituent des formes de négociation entre diverses instances sur des normes à mettre en place et favorisent la circulation de concepts (Deibert et Crete-Nishihata, 2012 ; Sire, 2017). Ces recommandations donnent à voir des techno-imaginaires sur l'éthique de l'IA, ouvrent des possibilités et en ferment d'autres, du fait des idéologies qu'elles propagent. L'AFC montre d'ailleurs que les instances s'influencent les unes les autres. Ainsi les associations, les publications multipartites, le secteur privé, tout comme les publications, dont les instances ne sont pas connues, utilisent les mêmes formes, ce qui suppose des représentations et univers de référence proches. Ce phénomène souligne un dialogisme entre les instances et donc des reprises ou négociations pour définir les normes de l'IA éthique. Les instances d'autorité publique et le secteur académique ont leurs propres formes pour s'exprimer sur le sujet et l'univers de référence des instituts de recherche semble le plus éloigné des autres catégories d'analyse.

En outre, avec des principes éthiques établis majoritairement par les pays du Nord, le risque de produire des formes de néo-colonialisme avec l'IA et ses données, mais aussi son éthique est élevé (Mohamed et al, 2020). Notamment parce que les sujets éthiques nécessitent de considérer les contextes culturels et géographiques[4] (Wong, 2020). Divers travaux alertent sur une approche globale de l'éthique de l'IA et signalent que celle-ci devrait être appréhendée de manière à travailler

4 « Moral Machine ». Consulté le 9 mars 2021. http://moralmachine.mit.edu. Cette expérience a pour objectif de récolter des données sur des dilemmes moraux.

avec et dans les cultures, afin d'établir collectivement des standards, en prenant en considération diverses formes d'existence. Tout en s'interrogeant sur la manière dont localement, les standards peuvent être appliqués en accord avec les valeurs du pays ou en montrant comment certaines normes, si nécessaires, seront bénéfiques dans le contexte local (Wong, 2020). L'éthique de l'IA devrait être « pensée avec » et non « pensée pour » afin d'éviter de (nouvelles) formes d'hégémonie, de domination ou de colonialisme avec la Technique, et ne pas imposer des « Systèmes Globaux » sous couvert d'objectivité (Haraway, 2007 et 2020).

En termes de contenu, l'analyse des similitudes montre que l'éthique n'est pas le sujet le plus assimilé à l'IA, malgré les fortes occurrences de « ethics » ou « ethical ». Ce lexique ne représente pas un halo à lui seul. L'éthique n'est pas une thématique principale et les définitions de l'IA restent technicistes. Son éthique comprend de grands concepts consensuels, mais aux applications parfois floues. Une définition vague ne peut que faire consensus entre les diverses instances, mais ne permet pas l'application de normes ou pratiques claires. La même remarque peut être faite pour le principe de « fairness », concepts qui renvoient à l'idée de produire des IA garantissant, si ce n'est une justice sociale (Fraser, 2004), au moins une absence de discrimination envers les personnes minorisées.

Nous ne nions pas que les instances œuvrent à l'établissement de principes sur l'éthique de l'IA, la prolifération des discours sur le sujet en témoigne. Toutefois, l'éthique ne semble pas être centrale et ne forme même pas un monde lexical à part dans le dendrogramme de Reinert. Sans compter que l'appellation « IA éthique » varie selon le domaine d'application ou le point de vue de l'instance, comme le montre le calcul des spécificités pour chacune des catégories. Par exemple, le secteur privé s'exprime à la première personne du pluriel « we » et au pronom possessif « our », ce qui témoigne de principes éthiques qui leur sont propres. Sans compter que le suremploi des termes « customer » ou « user » met en visibilité une cible restreinte et une éthique dédiées à celles.ceux qui consomment leur Technique. Pourtant, l'éthique n'a-t-elle pas vocation à protéger toutes les populations qu'elles soient clientes ou non ? Le secteur académique utilise les pronoms « you » et « your » et

s'adresse donc aux structures qui développent l'IA, renvoyant à une démarche de sensibilisation, tout en mettant en avant la transparence, l'explicabilité et la justice, spécifiquement en référence au terme « right(s) » associé aux droits humains. Quant aux institutions d'autorité publique, elles affichent un discours plutôt tourné vers les données, l'investissement et le développement d'une stratégie autour de l'IA.

Chacune des catégories semble adresser des problématiques et enjeux qui lui sont propres, et fait concorder ou décline l'éthique en fonction de ces derniers, avec : « a 'Pick Your Own 'approach to the identification of ethical standards for AI systems » (Yeung et al., 2019, 22). Il s'agirait alors d'une éthique de convenance, voire d'une sélection des lois auxquelles l'IA a pourtant obligation de se plier (Wagner, 2018). Ceci n'est pas sans rappeler un certain *ethic washing*, voire *social washing*, notamment parce que ces discours semblent refléter des obligations juridiques en matière de traitement des données ou de non-discrimination, bien plus que de proposer une nouvelle perspective éthique. De même, le principe de *fairness* comprend des terminologies très disparates qui renvoient à différentes notions de justice sociale (Fraser, 2004), à la fois au niveau sociologique, mais aussi juridique ou même dans l'application mathématique de ce principe lors du développement de l'IA (Verma et Rubin, 2018). Ainsi, les termes renvoient aux enjeux de reconnaissance, tels que : « diversity », tandis que certains sont liés à la reproduction de rapports de pouvoir comme « discriminatory » ou « bias », alors que d'autres reportent à différentes formes de justice sociale comme : « fairness » ou « equality ». Définir une justice sociale ne peut s'inscrire dans des principes flous, celle-ci nécessite une réflexion sociopolitique collective sur des modes de vie future. Toutefois, en restant dans_une définition technique de l'éthique et de la justice sociale, on est renvoyé à un imaginaire technicien et un solutionnisme numérique. Ainsi avant de concevoir de la Technique dite intelligente susceptible d'automatiser les rapports sociaux de sexe, de race et de classe et de chercher des solutions techniques à des problèmes sociétaux, l'éthique et la justice sociale ne devraient-elles pas venir poser des limites à ne pas dépasser ? Des définitions et objectifs qui s'inscrivent dans un projet de société ? Une approche « fair » ou éthique de l'IA devrait questionner les moyens qui assureront qu'il n'y ait pas d'oppressions algorithmiques pour les personnes à l'intersection de plusieurs formes de marginalisation, tout

en prenant en compte la multiplicité des contextes et vécus (Xiang et Raji, 2019).

Conclusion

Cet article visait à dépasser l'idée que l'IA serait neutre et objective en montrant qu'elle est imbriquée dans le monde social. Elle entre en relation avec les groupes sociaux qui le composent, produisant une co-construction entre IA et rapports sociaux de sexe, de race et de classe. L'approche intersectionnelle de l'IA a pour objectif de penser les enjeux de justice sociale pour les groupes minorisés à l'intersection de plusieurs formes de domination et de marginalisation. A partir d'un corpus de *soft law* sur la gouvernance éthique de l'IA, nous proposons une cartographie et analyse textométrique, qui décrivent les modes de circulation des principes de l'IA éthique et les traits saillants du corpus. L'accès à ces données ouvre un espace de réflexion et de critique sur l'éthique et le principe de « fairness » appliqué à l'IA en montrant les limites de l'imaginaire technicien pour dépasser les rapports sociaux, formes de domination et de marginalisation.

Toutefois, il s'agit d'une première entrée par la textométrie dans un vaste corpus, elle ne peut rendre compte intégralement des techno-imaginaires qui sont réinventés à travers ces discours sur l'éthique de l'IA. De plus, pour que l'analyse s'appuie pleinement sur une approche intersectionnelle, pas seulement en tant que concept, mais aussi en tant que méthode, une analyse sémantique des discours et plus spécifiquement de l'éthique et du principe de « fairness » est nécessaire, notamment pour comprendre comment ces recommandations transforment, actualisent ou réitèrent les rapports sociaux de sexe, de race et de classe, leur appréhension, leur compréhension et leur traitement de/par l'IA.

Bibliographie

Akrich, M. (1994). *Comment sortir de la dichotomie technique/société*. La Découverte. https://www.cairn.info/de-la-prehistoire-aux-missiles-balistiques--9782707123879-page-103.htm

Bakhtin, M. M. (1970). *Problèmes de la poètique de Dostoïevski*, L'Age d'Homme, Lausanne.

Baya-Laffite, N., Beaude, B., & Garrigues, J. (2018). Le deep learning au service de la prédiction de l'orientation sexuelle dans l'espace public. *Reseaux*, n° 211(5), 137–172.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*, Polity, Medford.

Berleur, J., & Poullet, Y. (2002). *Réguler Internet*. Etudes, Tome 397(11), 463–475.

Bernheim, A., Vincent, F., & Villani, C. (2019). *L'intelligence artificielle, pas sans elles !*, Editions Belin : Humensis, Paris.

Bertail, P., Bounie, D., Clémençon, S., & Waelbroeck, P. (2019). *Algorithmes : Biais, discrimination et équité*. Fondation ABEONA.

Bilge, S. (2009). Théorisations féministes de l'intersectionnalité. *Diogène*, 225(1), 70. https://doi.org/10.3917/dio.225.0070

Bilge, S., & Roy, O. (2010). La discrimination intersectionnelle : La naissance et le développement d'un concept et les paradoxes de sa mise en application en droit antidiscriminatoire*. Canadian Journal of Law and Society* / La Revue Canadienne Droit et Société, 25(1), 51–74. https://doi.org/10.1017/S0829320100010218

Cardon, D. (2013). *Politique des algorithmes : Les métriques du web*, La Découverte, Paris.

Cardon, D. (2015). *A quoi rêvent les algorithmes : Nos vies à l'heure des big data*, La République des idées : Seuil, Paris.

Chabaud-Rychter, D., & Gardey, D. (Eds.). (2002). *L'engendrement des choses : Des hommes, des femmes et des techniques*, EAC, Paris.

Cockburn, C. (1981). The Material of Male Power. *Feminist Review*, 9, 41. https://doi.org/10.2307/1394914

Collet, I. (2019). *Les oubliées du numérique : L'absence des femmes dans le monde digital n'est pas une fatalité*, Le Passeur éditeur, Paris.

Criado-Perez, C. (2019*). Invisible women: Exposing data bias in a world designed for men*, Vintage, London.

Datta, A., Tschantz, M. C., & Datta, A. (2015). *Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination.* Proceedings on Privacy Enhancing Technologies, 2015(1), 92–112. https://doi.org/10.1515/popets-2015-0007

Deibert, R. J., & Crete-Nishihata, M. (2012). Global Governance and the Spread of Cyberspace Controls. *Global Governance*, 18(3), 339–361.

Dorlin, E. (2008). *Sexe, genre et sexualités : Introduction à la théorie féministe (1. éd)*, Presses Univ. de France, Paris.

Dorlin, E., & Bidet-Mordrel, A. (Eds.). (2009). *Sexe, race, classe : Pour une épistémologie de la domination*, Presses Universitaires de France, Paris.

Doueihi, M. (2013). *Qu'est-ce que le numérique ?* Presses Universitaires de France. https://doi.org/10.3917/puf.doue.2013.01

Ellul, J. (1990). *La technique : Ou, L'enjeu du siècle* (2e éd. rev.), Economica, Paris.

Foucault, M. (2001). *Dits et écrits. 2: 1976 - 1988*, Gallimard, Paris.

Hamilton, I. A. (2018, October 10). *Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women*. Business Insider France. https://www.businessinsider.fr/us/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10

Hampton, L. M. (2021). Black Feminist Musings on Algorithmic Oppression. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 1–1. https://doi.org/10.1145/3442188.3445929

Haraway, D. J., Allard, L., & Gardey, D. (2007). *Manifeste cyborg et autres essais : Sciences, fictions, féminismes*, Exils, Paris.

Haraway, D. J., & García, V. (2020). *Vivre avec le trouble*, Les Editions des Mondes à faire, Vaulx-en-Velin.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Kergoat, D. (2012). *Se battre, disent-elles...*, La Dispute, Paris.

Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (n.d.). *How We Analyzed the COMPAS Recidivism Algorithm*. ProPublica. Retrieved June 9, 2020, from https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Latour, B., & Biezunski, M. (2010). *La science en action : Introduction à la sociologie des sciences*, Découverte / Poche, Paris.

Lavergne, B. (2013). *Recherche sur la soft law en droit public français*. Presses de l'Université Toulouse 1 Capitole. https://doi.org/10.4000/books.putc.1866

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, 13(5), 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Maingueneau, D. (2014). *Discours et analyse du discours : Une introduction*, Armand Colin, Paris.

Mazabraud, B. (2010). Foucault, le droit et les dispositifs de pouvoir. *Cites*, n° 42(2), 127–189.

Mazouz, S. (2020). *Race*, Anamosa, Paris.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00405-8

Musso, P. (2020). Le désir technologique de Dieu. *Quaderni*, n° 99-100(1), 113–124.

Née, É. (Ed.). (2017). *Méthodes et outils informatiques pour l'analyse des discours*, Presses universitaires de Rennes, Rennes.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*, New York University Press, New york.

Reinert, M. (1993). Les « mondes lexicaux » et leur « logique » à travers l'analyse statistique d'un corpus de récits de cauchemars. *Langage & société*, 66(1), 5–39. https://doi.org/10.3406/lsoc.1993.2632

Saigh, D., Borzic, B., Alkhouli, A., & Longhi, J. (2017). Contribution linguistique à une classification automatique des communautés de sens et à leur analyse. *Questions de communication*, n° 31(1), 161–182.

Sire, G. (2017). Gouverner le HTML. *Reseaux*, n° 206(6), 37–60.

Tatman, R. (2017). *Proceedings of the First Workshop on Ethics in Natural Language Processing*. Association for Computational Linguistics, 53–59.

van Zoonen, L. (1992). Feminist theory and information technology. *Media, Culture & Society*, 14(1), 9–29. https://doi.org/10.1177/016344392014001002

Verma, S., & Rubin, J. (2018). Fairness definitions explained. *Proceedings of the International Workshop on Software Fairness*, 1–7. https://doi.org/10.1145/3194770.3194776

Wagner, B. (2018). Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping? *In BEING PROFILED:COGITAS ERGO SUM* (pp. 84–89). Amsterdam University Press. https://www.degruyter.com/document/doi/10.1515/9789048550180-016/html

Winner, L. (1980). *Do Artifacts Have Politics?* The MIT Press, 109(1, Modern Technology: Problem or Opportunity?), 121–136.

Wong, P.-H. (2020). Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00413-8

Xiang, A., & Raji, I. D. (2019). *On the Legal Compatibility of Fairness Definitions*. ArXiv :1912.00761 [Cs, Stat]. http://arxiv.org/abs/1912.00761

Yeung, K., Andrew Howes, & Pogrebna, G. (2019). AI Governance by Human Rights-Centred Design, Deliberation and Oversight: An End to Ethics Washing (SSRN Scholarly Paper ID 3435011). Social Science Research Network. https://doi.org/10.2139/ssrn.3435011