SHAPING AI

Sciences et dissonances : les discours multiples du secteur pancanadien de la recherche en IA (2012-2023)

Nicolas Chartier-Edwards
Etienne Grenier
Robert Marinov
Guillaume Dandurand
Fenwick McKelvey
Jonathan Roberge

INRS Institut national de la recherche scientifique

À propos de ce document

Le contenu de ce rapport est issu des activités de la division canadienne du consortium de recherche international Shaping AI (2021 - 2024) sous la direction du professeur Jonathan Roberge. L'équipe de recherche tient à remercier les représentant.es de la communauté scientifique qui ont accepté de partager leurs expériences à propos de l'IA, telle qu'elle se conçoit au Canada.

Ce rapport s'appuie sur des recherches financées par le Conseil de recherches en sciences humaines.

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada

ISBN : 978-2-89575-476-3

Équipe de recherche

Shaping AI est un projet de recherche social multinational et multidisciplinaire consacré aux trajectoires globales des discours publics visant l'intelligence artificielle (IA) dans quatre pays (Allemagne, Royaume-Uni, Canada, France) sur période de douze ans (2012 – 2024).

L'équipe de recherche canadienne est dirigée par :

Jonathan Roberge est professeur titulaire à l'Institut national de la recherche scientifique (INRS) à Montréal. Il est le fondateur du Nenic Lab (Chaire de recherche du Canada sur les nouveaux environnements numériques et l'intermédiation culturelle) et est cotitulaire de la Chaire de recherche du Québec sur l'intelligence artificielle et le numérique francophones. Parmi ses plus récents ouvrages édités figurent *Algorithmic Culture* (Routledge, 2016) et *The Cultural Life of Machine Learning* (Palgrave, 2020).

Fenwick McKelvey est professeur agrégé à l'Université Concordia et se spécialise dans le domaine des politiques technologiques de l'information et de la communication. Il est l'auteur de *Internet Daemons* (University of Minnesota Press, 2018) et gagnant de l'édition 2019 du prix Gertrude J. Robinson Book décerné par l'Association canadienne de communication.

Pour ce rapport, l'équipe de recherche inclut également :

Nicolas Chartier-Edwards est étudiant au Doctorat sur mesure à l'Institut national de la recherche scientifique (INRS) et détient une maitrise en sociologie de l'Université Laval. Ses travaux portent principalement sur l'intégration des technologies d'intelligence artificielle dans la production du social et plus précisément, dans le cadre de sa thèse, sur la transformation de la gouvernance canadienne par le déploiement de ces mêmes technologies dans les fonctions administratives des différents paliers étatiques.

Etienne Grenier est un artiste et chercheur évoluant dans le domaine des cultures numériques. Inscrit au doctorat à l'Institut national de la recherche scientifique, il étudie les impacts de la datatification sur les chaînes de production culturelles et contribue au projet Shaping AI. Sa pratique créative dans le domaine des arts numériques l'a conduit à diffuser ses œuvres au sein d'institutions de premier plan et de festivals majeurs en Europe et dans les Amériques.

Robert Marinov est un étudiant inscrit au doctorat à l'université Concordia. Il possède une maîtrise en science politique obtenue à l'Université d'Ottawa, où son mémoire lui attribua le prix de la Commission des études supérieures en humanités. Ses activités de recherche se concentrent sur les plateformes émergentes de jumeaux numériques ainsi que sur leur intersection avec l'intelligence artificielle sous l'angle des enjeux de gouvernance et d'impacts environnementaux qu'elles génèrent. Il a publié ses travaux dans plusieurs revues scientifiques, notamment *Canadian Journal of Political Science, Communication Review, Politics & Policy, et Critical Studies in Media Communication*.

Guillaume Dandurand est un chercheur postdoctoral associé à l'Institut national de la recherche scientifique, à Montréal. Il est le coéditeur de *Les économies de la promesse* (Presses de l'Université de Montréal, 2022). Ses travaux de recherche au doctorat lui ont permis d'obtenir le Dissertation Prize, attribué par l'université York (2019) et de l'inclure dans la sélection de la compétition pour la dissertation en sciences sociales organisée par l'International Convention of Asia Scholars (2021).

FAITS SAILLANTS

- Le Canada offre un contexte particulier pour la recherche en AI compte tenu entre autres de la présence de trois instituts de recherche fortement encouragés par la Stratégie pan-canadienne et le CIFAR. Ceci étant, il existe nombre d'interprétations, de divergences et de controverses quant au développement de l'IA.

- Une crise de sens affecte la notion même d'intelligence artificielle créant ainsi une forme de confusion qui affecte le monde de la recherche et les pratiques des experts.

- Des enjeux d'accès à des ressources financières, computationnelles et informationnelles suffisantes dans un paradigme dominé par les grands modèles d'IA font en sorte que l'écosystème canadien de la recherche fait face à des défis en ce qui a trait à son positionnement et ses performances et ce, principalement dans un environnement marqué par une science industrialisée d'origine américaine.

- Le battage médiatique entourant le déploiement de l'IA et le faible niveau de littératie dans la population générale participent à la mise en place d'un contexte où la gouvernance de l'IA échappe partiellement au contrôle démocratique.

- Il existe un important décalage entre le récit promotionnel de l'IA promulgué par des institutions étatiques et médiatiques canadiennes et la situation réelle des chercheurs appartenant à cet écosystème.

- L'enjeu de la désinformation est un des rares à faire l'unanimité, du moins en surface.

TABLE DES MATIÈRES

1. Introduction

Problématiser la recherche actuelle portant sur l'IA au Canada équivaut à examiner une vaste gamme de récits dont les répercussions sont tangibles. Aujourd'hui, lorsque les acteurs, les institutions et les organisations font référence à l'IA, ils décrivent la mesure dans laquelle l'IA fait déjà partie intégrante du tissu social, économique et politique. Par exemple, le ministère de l'Innovation, des Sciences et du Développement économique du Canada a déclaré que « l'IA [...] a déjà un impact important sur la vie quotidienne des Canadiens » (Innovation, Sciences et Développement économique Canada, 2023). Au provincial, Pierre Fitzgibbon, ministre de l'Économie, de l'Innovation et de l'Énergie sous le gouvernement de la Coalition Avenir Québec, a récemment déclaré que « l'IA touche progressivement tous les secteurs de la société, ce qui implique que des balises soient considérées lors de son développement et de son utilisation » (Gouvernement du Québec, 2023). McKinsey and Company[1], le CIFAR et Scale AI suggèrent également que l'IA fait désormais partie intégrante de toutes les sphères d'activité au pays. Comment cet imaginaire s'est-il progressivement mis en place pour, encore une fois, participer du déploiement bien réel de la technologie ? Et, d'abord, que peuvent en dire et en penser les personnes mêmes qui s'y impliquent le plus directement, à savoir les scientifiques œuvrant dans différents laboratoires, allant des sciences computationnelles jusqu'aux sciences sociales et à la philosophie ? La présente étude s'intéresse principalement à ces praticiens et praticiennes ainsi qu'aux différentes manières dont ils et elles conçoivent ce qu'est et fait l'IA aujourd'hui. Il est de ce fait question du type de production de connaissance émanant des laboratoires, mais également de la compréhension de ce qui est plus largement en jeu, à savoir les valeurs qui sont mobilisées pour faire sens de ce qui se transforme dans et par l'IA. Ce qui est ainsi présenté est une gamme relativement complexe de positions le plus souvent adaptatives ; ce qui dans l'étude prend la forme d'un arc entre ce qui est plus « computo-centrique » et ce qui est davantage « socio- centrique » (Marres et coll., 2024). Cette gamme – ou arc – existe parce que le champ technoscientifique de l'IA est aujourd'hui fortement marqué par ce que Pinch et Bijker nomment canoniquement sa « flexibilité interprétative » (1984). Pour ces derniers, en effet, « studying technologies as objects that are 'socially constructed' involve recognizing how their form and functions

[1] L'état de l'IA en 2023 : L'année de rupture de l'IA générative » (McKinsey & Co, 20 novembre 2023).

come out of the process of interpretative negotiation among producers, sellers, and other relevant social actors » (Magaudda, 2014, p. 66). Qui négocie quoi et comment dans le champ de la recherche en IA au Canada ? De quelle manière cela oriente-t-il ledit champ et plus largement la société ? Ce sont ces questions qui occupent le présent rapport.

En 2024, l'étude de l'IA présente des caractéristiques uniques. Tout d'abord, la technologie est toujours un domaine de recherche en plein essor. Le nombre de scientifiques s'y consacrant a explosé tant au pays qu'à l'international, les publications dans les diverses revues spécialisées ont elle aussi connu une croissance exponentielle, et ce, au même moment où la compétitivité pour accéder à des conférences telles que NeurIPS ou ICML a énormément crue. De même, les panels sur l'IA dans les conférences consacrées à l'étude des sciences et des techniques comme 4S ou encore EASST sont maintenant légion alors que ce n'était pas le cas il y a encore quelques années. Deuxièmement, le domaine progresse très rapidement. Comme il en sera largement fait mention dans le rapport, le lancement de ChatGPT a ébranlé des communautés entières de recherche en IA : des chercheurs en sciences sociales qui ont rapidement porté leur attention sur l'IA générative comme des informaticiens qui utilisent les plateformes de médias sociaux et traditionnels pour débattre, par exemple, des dangers existentiels de l'IA pour la société, sinon, de la civilisation elle-même. Troisièmement, le rythme de la recherche est si rapide qu'il modifie les pratiques mêmes des laboratoires. Les informaticiens qui avaient l'habitude de fonder la légitimité de leurs travaux sur des recherches évaluées par des pairs ne peuvent plus se permettre de les faire passer par ce même processus afin de rester « à jour » avec les plus récentes avancées. Comme il s'agira de le voir, les travaux sur l'apprentissage automatique sont de plus en plus évalués en fonction de la quantité de paramètres utilisés et de leurs capacités à atteindre des performances dites « state-of-the-art » (ou SOTA) (Lipton et Steinhardt, 2018). Quatrièmement, la recherche en IA n'est plus confinée qu'aux espaces académiques. Par exemple, les projets financés par des fonds publics se déroulent désormais dans un environnement où les entreprises privées sont devenues des acteurs clés – soit via un partenariat public-privé ou à travers la cooptation de chercheurs universitaires par des plateformes et des entreprises comme Google, Uber, etc. (Roberge et coll., 2019). Cinquièmement, comme le suggèrent Marres et Gerlitz (2015), les notions de conscience sociale quant aux impacts de la technologie – son « éthique » – font leur chemin dans la recherche en sciences computationnelles, façonnant au passage ce qu'est l'IA et ce qu'elle pourrait devenir. Comme il en sera amplement fait mention dans ce rapport, une caractéristique importante du champ apparait comme étant son ambiguïté même. Ici, l'IA est une discipline caractérisée par des couches superposées

de classificateurs linéaires, de calculateurs sans cesse plus puissants et de progrès technologiques très rapides. Maintenant, elle est pensée comme s'apprêtant à envahir le monde. Parfois, sinon souvent, ces différentes versions, interprétations et possibilités s'entremêlent. Dans le présent rapport, c'est cette nature ambiguë de l'IA qui est mise à contribution pour étudier comment les chercheurs canadiens donnent un sens à leur objet de recherche et à sa place en société.

Ensemble, ces caractéristiques soulèvent une série de questions sur la trajectoire de la recherche en IA au pays et sur la manière dont un objet aussi polyvalent a pu, à ce jour, mobiliser un tel niveau de ressources. Pour étudier cet état de la recherche en IA, le rapport se tourne principalement vers les chercheur.es – experts en science computationnelle, de la santé et des sciences sociales – qui ont une expertise contributive sur le sujet (Collins et Evans, 2002). L'objectif est d'examiner de manière critique – comme cela est fait plus en détail dans les chapitres suivants cette introduction –, la manière dont ces spécialistes perçoivent la situation de la recherche en IA au pays et la manière dont leurs travaux et objets de recherche façonnent à leur tour les compréhensions collectives et concurrentes de l'IA – ce qui est autrement nommé la *négociation sociale* de l'IA.

1.1 Contexte

1.1.1 L'évolution de la recherche en IA au Canada

Après plus d'une décennie de présentation et de représentation médiatique, l'histoire de l'intelligence artificielle est devenue chose plus familière. Ponctuée de controverses scientifiques, sa trajectoire a connu une série de printemps et d'hivers au cours desquels autant l'intérêt envers l'IA et son financement ont radicalement changé (Cardon et coll., 2018 ; Roberge et Castelle, 2021). Ce que l'on appelle communément « l'IA » aujourd'hui trouve son origine dans les années quarante avec entre autres le *Perceptron*, un réseau neuronal conçu pour reconnaître des images sur la base d'un classificateur linéaire attribuant des poids statistiques à des valeurs pour ensuite identifier la catégorie à laquelle elles appartiennent (Hunt et Lepage-Richer, 2024). À cette époque, le concept de machine pensante a inspiré plusieurs projets scientifiques à la fois connexionnistes et davantage symboliques (Cardon et coll., 2018). De fait, aucun n'a reçu autant d'attention que la conférence de John McCarthy et de ses collègues du collège de Dartmouth proposant d'étudier l'idée ou le concept même d'« intelligence artificielle » (1955, p. 1). Contrairement aux techniques de réseaux neuronaux utilisant le cerveau comme source

d'inspiration matérielle pour la conception d'algorithmes, une IA plus symbolique – et pour ainsi dire abstraite – aspire, selon les termes de McCarthy et de ses collègues, à « simuler [la] caractéristique de l'intelligence » (1955, p. 1). Depuis lors, l'IA est passée par une série de paradigmes – le cognitivisme informatique et les systèmes experts, par exemple – pour en arriver au tournant des années 2010, moment où les progrès technologiques en matière de puissance de calcul, associés à l'accès récent aux médias sociaux et à des ensembles volumineux de données (Big Data), ont fait place aux récents succès de l'apprentissage automatique (Mendon-Plasek, 2021). Suite aux démonstrations récentes de ses performances, c'est bien ainsi cette IA dite « connexionniste » propulsée par le *deep learning* qui a gagné en popularité en dehors des laboratoires, et ce, au point d'être qualifiée de « révolutionnaire », sinon de « dangereuse » par les mêmes experts en l'espace de quelques années (Bengio, 2016 ; 2023).

Cette histoire familière – celle d'une ascension, d'une chute et d'une remontée – a façonné la compréhension populaire de l'IA et l'a présentée comme un objet qui, jusqu'à récemment, n'était que mathématique, théorique et conceptuel. Pourtant, ce récit manque de fondements épistémologiques et culturels. Tout comme les infrastructures hydroélectriques ou la machine à vapeur, l'apprentissage profond n'est pas une technique neutre (Winner, 1980). Imprégnée de significations et de politique, cette technique d'apprentissage est fermement située dans un contexte spatial et temporel : ici très justement au Canada où plusieurs des récents apports de la recherche en IA se sont historiquement effectués.

Au pays, comme partout ailleurs, l'IA est intimement liée à une compréhension des technologies comme moyen de propulser le pays à titre de chef de file mondial de l'innovation (Attard-Frost, 2022; McKelvey, 2018). Ce genre de conceptions font des systèmes d'IA des atouts dits nationaux créant un avantage économique et faisant du Canada un endroit où l'économie prospérerait (Birch, 2020). Mais ladite technologie ne se résume pas à cela. Au Canada, l'IA est également une source de fierté qui fait du pays une destination de choix pour nombre de chercheurs en réseaux neuronaux. Deux des informaticiens les plus accomplis dans le champ, Geoffrey Hinton et Yoshua Bengio, résident et travaillent au Canada. Les agences gouvernementales de financement ont d'ailleurs soutenu leurs travaux, et ce, même pendant les longues périodes où l'IA n'était pas un domaine d'étude à la mode. Aussi, plus récemment, c'est bien une pléthore d'acteurs qui se sont agglomérés autour d'eux pour créer un environnement politique propice à l'adoption rapide de l'IA par les industries locales (Roberge et coll., 2019).

Cette matrice épistémologique et culturelle remonte à la guerre froide, lorsque feu le premier ministre canadien Pierre-Elliott Trudeau (1968-79 ; 1980-84) a affirmé que les technologies émergentes de la communication et des médias étaient essentielles, car pouvant « transformer toute notre société » (1969, dans Lepage-Richer et McKelvey 2022, p.7). Ces mêmes technologies devaient fournir les moyens de réorganiser la bureaucratie et la gestion de l'information afin de favoriser une administration plus efficace des affaires de l'État. Avec l'émergence des systèmes experts, des formes plus accessibles de puissance informatique ont offert de nouvelles promesses d'opportunités économiques et de nouveaux moyens. Ainsi, au début des années 1980, les premières itérations de l'IA étaient considérées comme des technologies susceptibles de changer radicalement la société. En guise d'exemple, la *1982-84 Royal Commission on the Economic Union and Development* (RCEUD) émit comme recommandation que l'État soutienne la recherche sur l'IA puisqu'offrant la possibilité de favoriser « the development of a coast-to-coast pool of expertise in the field as another opportunity to foster economic development and national unity alike » (2022).

Pour déterminer comment tirer le meilleur parti des avantages sociaux et économiques de l'IA, la RCEUD mobilisa l'expertise de l'Institut canadien de recherches avancées (CIFAR). Aujourd'hui, le CIFAR est connu comme l'institution clé ayant soutenu la recherche portant sur l'IA, et ce, même durant ce que la communauté des informaticiens appelle l'hiver de l'IA. Au Canada, le CIFAR est crédité, selon Robert Hunt et Théo Lepage Richer, « with having almost single-handedly catalyzed the rise of 'neural networks' » (Hunt & Lepage-Richer, 2024). Le CIFAR contribue encore aujourd'hui à ce récit en soulignant sur son propre site web comment l'institution a « kickstart[ed] the revolution in artificial intelligence that powers the modern world[2]. » Toutefois, pour Hunt et Lepage-Richer (2024), la principale contribution du CIFAR ne réside pas tant dans les succès des réseaux neuronaux que dans la cristallisation des réseaux de recherche en tant que modèle dominant pour la mise en œuvre des priorités de recherche au pays. Dès le début des années 1980 en effet, le CIFAR a présenté un modèle de gouvernance par priorités de recherche s'alignant sur celui de la RCEUD. Il a ainsi permis la création de passerelles entre les universités, les entreprises et les niveaux de gouvernement encourageant une approche décentralisée de la recherche autour d'objets technoscientifiques préidentifiés, et ce, grâce à l'allocation de fonds répondant directement aux priorités de recherche de l'État et du CIFAR.

[2] Accessible à l'adresse suivante : https://cifar.ca/our-story/ (dernier accès le 03 octobre 2023).

Selon son fondateur John Leyerle, le mandat initial du CIFAR était de « foster basic, conceptual research of high quality at an advanced level across the full spectrum of knowledge » (1979, cité sur le site web du CIFAR)[3]. Le CIFAR est une organisation caritative enregistrée qui reçoit des fonds de trois organismes gouvernementaux : l'État fédéral et les provinces de l'Alberta et du Québec) ainsi que d'entreprises, de fondations, de particuliers et d'organisations situés à l'intérieur et à l'extérieur du Canada. Une telle répartition des fonds destinés à la recherche universitaire est donc politique à la fois par nature et par conception. Elle oriente les priorités de recherche en stabilisant les réseaux de recherche tout en empêchant leur examen public (Hunt et Lepage-Richer, 2024).

Suite à l'identification de l'IA par le comité scientifique du CIFAR comme un programme de recherche clé au début des années 1980, le même organisme a offert un poste à Geoffrey Hinton à l'Université de Toronto afin de participer au premier programme du *CIFAR, Artificial Intelligence, Robotics & Society*. Les comptes rendus de cette période et des années subséquentes suivent généralement un récit similaire : i) beaucoup pensaient que la recherche sur les réseaux neuronaux était une perte de temps et d'argent ; ii) mais grâce aux dotations financières du CIFAR pour poursuivre la recherche sur des « idées scientifiques non orthodoxes » (Onstad, 2018), les chercheurs canadiens ont pu contre toute attente développer des techniques de réseaux neuronaux, ; iii) éventuellement, d'autres chercheurs commencèrent à graviter autour de ces mêmes idées dites « non orthodoxes » ; iv) l'émergence de la puissance de calcul et des données massives fut tributaire d'un gain en performance du côté des techniques de réseaux neuronaux, permettant par le fait même à ces dernières d'éclipser l'école de l'IA symbolique ; v) ce qui engendra un gonflement, sinon une explosion des représentations et des applications futures de l'IA (Cardon et coll., 2018 ; Dandurand et coll., 2022 ; Onstad, 2018). « We were outsiders, but we also felt like we had a rare insight, like we were special », déclare Ilya Sutskever, cofondateur d'OpenAI, se souvenant de ses années de formation dans le laboratoire de Hinton au début des années 2000. « We were clearly outside the establishment […]. It's funny: now we've become the establishment », note-t-il encore. Le CIFAR aime par ailleurs rappeler aux personnes qui naviguent sur ses pages web qu'il a eu la clairvoyance d'encourager et de créer « un nouvel âge de l'IA » alors que personne d'autre n'y croyait[4].

[3] Accessible à l'adresse suivante : https://cifar.ca/our-story/ (dernier accès le 03 octobre 2023).
[4] Accessible à l'adresse suivante : https://cifar.ca/our-story/ (dernier accès le 03 octobre 2023).

Le programme *Intelligence artificielle, robotique et société* fut fermé en 1995, mais Hinton est resté actif au sein du CIFAR. Au début des années 2000, il a participé à la création et est devenu le premier directeur du programme *Neural Computation & Adaptative Perception*, finalement devenu le programme *Learning in Machines and Brains* en 2014 et qui comprenait des experts tels que Yoshua Bengio et Yan LeCun (Senneville, 2021). Hinton a quitté la direction du programme en 2013 lorsque Google a racheté sa startup, DNNresearch. Il a ensuite été consultant pendant une décennie jusqu'à ce qu'il démissionne en estimant que l'IA générative pose des « risques existentiels » (Taylor et Hern, 2023)[5]. En 2019, Bengio et LeCun étaient tous deux codirecteurs du programme[6], devenant simultanément, avec Hinton, mondialement reconnus en tant que lauréats du prestigieux prix Turing 2018, formant ce que certains qualifient – de manière plus caustique – la mafia canadienne de l'IA (Bergen et Wagner, 2015).

Plus précisément, c'est l'année 2012 qui marqua le début du dernier et actuel cycle d'engouement pour l'IA. En utilisant des techniques d'apprentissage profond, l'équipe de Hinton remporte le concours ImageNet avec une marge supérieure à celles d'autres méthodes (Cardon et coll., 2018). La démonstration est ainsi faite au reste du monde qu'avec puissance de calcul et disponibilité de données volumineuses, les techniques de réseaux neuronaux s'avèrent sans conteste la plus efficace. La communauté scientifique s'intéresse dès lors à l'IA dite « connexionniste » et en vient à formuler de nouvelles promesses d'ordre technoscientifiques (Borup et coll., 2006 ; Dandurand et coll., 2022). Au Canada, les intérêts privés s'intéressent eux aussi de plus en plus aux possibilités offertes par les techniques de réseaux neuronaux. Rapidement, les entreprises commencent à se regrouper autour d'informaticiens pour traduire la recherche sur l'apprentissage profond en applications pratiques et commercialisables. Comme mentionné plus haut, la startup de Hinton est rachetée par Google afin qu'Hinton puisse mettre son expertise au service de l'entreprise californienne dès 2013. En 2015, Bengio est nommé directeur scientifique de l'Institut de valorisation des données (IVADO), une organisation dont la mission est de « développer et promouvoir »[7] l'IA en plus de devenir consultant pour de nombreuses multinationales, dont Microsoft (Hempel, 2017)[8]. Lors d'un TedTalk X à Montréal l'année suivante, Bengio monte sur scène pour annoncer que l'IA allait engendrer la prochaine « révolution industrielle » (cité dans Colleret et Gingras, 2022). La même année, Bengio

[5] Le "parrain de l'IA" Geoffrey Hinton quitte Google et met en garde contre les dangers de la désinformation | Google | The Guardian

[6] Aujourd'hui, le programme est codirigé par Bengio et Konrad Kording, un neuroscientifique américain de l'université de Pennsylvanie.

[7] D'après le site web d'IVADO : A propos de nous | IVADO, consulté le 12 octobre 2023.

[8] Le retour de Microsoft sur le devant de la scène de l'intelligence artificielle | WIRED

cofonde une startup nommée Element AI. Dans les médias, la compagnie est présentée comme le moteur capable de mener à bien cette révolution ; une sorte de projet national promouvant l'idée qu'Element AI « pourrait » faire du Québec un leader mondial en IA (Roberge et coll. 2022). Cette dernière, à défaut de réussir à développer un produit et à trouver un marché, sera vendue en novembre 2020. Au total, Element AI aura reçu quatre tours de financement entre 2016 et 2020. Microsoft aura offert un premier tour d'amorçage par l'intermédiaire de sa filiale, M12. Quelques mois plus tard, en juin 2017, Element AI aura ensuite reçu 137,5 millions de dollars canadiens dans le cadre d'un financement de série A de la part de nombreuses sociétés de capital-risque[9]. À l'automne 2019, Element AI leva 200 millions CAD supplémentaires dans le cadre d'un cycle de série B auprès de nouveaux investisseurs publics et privés (Halin et Laroque, 2020 ; La Presse Canadienne, 2019)[10]. Finalement, Element AI a obtenu des fonds pour rembourser ses débiteurs (Roberge et coll., 2022) alors qu'elle a été vendue à la multinationale américaine ServiceNow pour 230 millions CAD, deux ans donc après que sa valorisation ait été estimée à 1 milliard CAD (George-Cosh, 2018 ; Silcoff, 2020 ; Scott, 2021).

Depuis 2012, les promesses technoscientifiques voulant que l'IA engendre une nouvelle révolution industrielle ont également convaincu la classe politique de créer les conditions optimales pour mettre en œuvre le développement et le déploiement de la technologie. En 2017, le gouvernement du Canada lança sa stratégie pancanadienne en matière d'IA afin de « favoriser l'adoption de l'intelligence artificielle dans l'ensemble de l'économie et de la société canadiennes » (Gouvernement du Canada, 2022). La stratégie repose sur trois piliers : la commercialisation, les normes et la recherche. En ce qui concerne la commercialisation, la stratégie a contribué à créer trois instituts nationaux d'IA responsable de traduire la recherche fondamentale en applications commercialisables. Ces trois instituts sont l'AMII à Edmonton (dirigé par Richard Sutton), le Vector Institute à Toronto (dirigé par Geoffrey Hinton) et le MILA à Montréal (dirigé par Yoshua Bengio). Ils ont pour mission non seulement de promouvoir l'IA, mais aussi de créer des connexions, souvent sous la forme de partenariats, avec d'autres organisations privées afin encore une fois de commercialiser l'IA. Dans ce cadre de la commercialisation, le Canada a également créé une « supergrappe » appelée *AI Global Innovation Cluster* qui fait quant à elle la promotion de l'adoption de l'IA dans les organisations, entreprises et industries canadiennes. Cette dernière est gérée de manière opaque par Scale AI à Montréal (Halin, 2024)[11]. Le deuxième

[9] Il s'agit de Data Collective Venture Capital, Real Ventures, la Banque de développement du Canada, Intel Capital, Microsoft Ventures, NVIDIA et Tencent.

[10] Il s'agit notamment de McKinsey & Company, de la Caisse de dépôt et placements et du gouvernement du Québec.

pilier renvoie quant à lui à la création et diffusion de normes relatives à l'IA. Par exemple, le gouvernement du Canada s'est récemment efforcé de créer un *guide sur l'utilisation de l'IA générative* dans les institutions publiques[12]. Finalement, le troisième pilier concerne la recherche et vise à attirer et à retenir les « talents » en matière d'IA au Canada. Par exemple, le gouvernement du Canada offre des chaires de recherche en IA octroyées et gérées par le biais du CIFAR pour stimuler la recherche dans des domaines particuliers ainsi que, mais dans une moindre mesure, les applications sociales de l'IA. Bien que le financement de ces chaires provienne directement du gouvernement du Canada, le CIFAR ne divulgue pas l'identité du comité scientifique qui les pilote ni les modalités exactes par lesquelles ces chaires sont distribuées.

Pour prendre cet autre exemple, entre 2017 et 2020, le gouvernement du Québec a investi 475 millions de dollars canadiens dans son écosystème afin de financer notamment la création de Forum IA Québec, un organisme sans but lucratif créé pour promouvoir l'adoption de l'IA ainsi qu'établir l'Observatoire international sur les impacts sociétaux de l'IA et du numérique (OBVIA). Selon le Fonds de recherche du Québec, OBVIA, dont le mandat est de servir d'« espace de discussion et de réflexion pour tous les intervenants concernés par l'IA »[13], est intimement lié à un réseau d'acteurs et d'organisations politiques, industrielles et universitaires qui travaillent à faire de l'IA une ressource clé pour l'économie québécoise. Son budget annuel est de 2,8 millions de dollars canadiens, une somme largement bonifiée récemment suite à un (non) – concours de MDIE. Il est important de noter que l'OBVIA ne se positionne pas comme le dernier rempart institutionnel contre l'engouement et le battage médiatique de l'IA ou encore, comme garde-fou contre les effets inattendus, voire délétères, pouvant être engendrés par cette technologie. Cet observatoire produit plutôt du matériel de recherche normalisant l'IA comme un objet qui « change l'organisation de notre monde », comme l'affirme son dernier rapport (Langlois et coll. 2023, 4).

Pour citer Maxime Colleret et Yves Gingras, un tel écosystème de l'IA québécois doit être compris comme étant « tissé serré » (2022 ; voir également Roberge, 2019). Les intérêts des universitaires, des décideurs, des commentateurs publics, des investisseurs en capital-risque et des industriels y sont dynamiquement « alignés ». En fait, comme le soulignent

[11] Depuis 2018, Scale AI a reçu 284 millions CAD du gouvernement du Canada et 53 millions du gouvernement du Québec, ainsi que plus de 250 millions CAD de financement supplémentaire non divulgué du secteur privé. https://www.scaleai.ca/fr/a-propos/, consulté le 30 septembre 2024.

[12] https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html, consulté le 12 octobre 2023.

[13] https://www.obvia.ca/a-propos, consulté 30 septembre 2024.

Jonathan Roberge et ses collègues, la mise en place systématique d'institutions et d'organisations qui partagent l'objectif de créer un tel écosystème repose sur un « schematic model of innovation where corporate actors come to mesh with university and government ones » (2019, p. 130 ; voir aussi Etzkowitz & Leidesdorff 2000 ; Dandurand et coll. 2022). Un tel modèle brouille souvent les frontières entre la recherche universitaire (financée par des fonds publics), les priorités politiques et les intérêts privés des industries. Il tend à justifier l'existence de portes tournantes entre la recherche universitaire et les industries par lesquelles les chercheurs et les investisseurs transigent sans trop d'efforts. Dans le discours public, les acteurs à l'avant-garde de ce réseau deviennent les porte-parole de l'adoption de l'IA (Callon 1980 ; 1984) et occupent de nombreuses positions de pouvoir dans l'écosystème. Par exemple, Bengio est *à la fois* professeur à l'Université de Montréal, directeur scientifique de Mila et d'IVADO, codirecteur du programme d'IA au CIFAR et consultant pour ServiceNow. Plus encore, ce dernier a notamment été codemandeur de la proposition de recherche à l'origine de l'OBVIA. Et peut-être plus important encore, il copréside le Conseil consultatif sur l'IA au sein d'Innovation, Sciences et Développement économique Canada, l'agence fédérale qui « conseille le gouvernement du Canada » sur toutes les questions liées à l'IA[14]. Joëlle Pineau, pour donner un exemple supplémentaire, « partage son temps » entre l'Université McGill et Meta Lab, à Montréal, où elle est directrice générale[15]. Quant à elle, Lyse Langlois est directrice générale de l'OBVIA et a siégé au conseil d'administration de Forum IA Québec. Avant même la création d'OBVIA en 2018, ces trois professeur.es ont conjointement fait partie de l'équipe scientifique à l'origine de la *Déclaration de Montréal pour une IA responsable*. Cette Déclaration a été conçue comme une boîte à outils dite « éthique » pouvant guider de manière non coercitive le développement de l'IA. Selon son propre site web, plus de 500 parties prenantes ont participé à un processus de consultation de trois mois aboutissant à sa création. Cependant, comme le montrent Roberge et ses collègues (2020), ce même exercice dit « délibératif et inclusif » a été en quelque sorte conçu pour exclure les voix critiques importantes dans les études sur l'IA et les mégadonnées. En d'autres termes, le comité scientifique d'une déclaration qui met l'accent sur les principes éthiques comme moyen de protéger la société contre les abus, le favoritisme et les actes répréhensibles – pour l'orienter plutôt vers une IA responsable – est largement composé d'informaticien.nes et d'autres chercheur.es et entrepreneur.es qui bénéficieraient directement d'un déploiement à grande échelle de l'IA dans la société. Aujourd'hui,

[14] Le Conseil consultatif fait des recommandations au ministre de l'Innovation, des Sciences et du Développement économique. De plus amples informations sont disponibles sur le site Conseil consultatif sur l'intelligence artificielle (canada.ca) (dernier accès le 18 octobre 2023).

[15] Disponible ici : https://www.cs.mcgill.ca/~jpineau/ (dernier accès le 18 octobre 2023).

l'expression « IA responsable » semble normalisée un peu partout au Canada, autant dans les institutions publiques que les organisations privées (voir Chartier-Edwards et coll., sous presse ; McKelbey et coll., 2024). En effet, l'appellation IA responsable reste vague et peut être trop facilement être détourné, dénaturé ou même vidé de son sens (Attard-Frost et Widder, 2023 ; Munn, 2023). Comme il s'agira de le voir à différents endroits de ce rapport, l'idée même d'une IA responsable participe d'une ambiguïté pour ainsi dire « performative ».

1.1.2 Position du rapport dans le vaste chantier Shaping 21st Century AI, Controversies and Closure in Media, Policy, and Research

Depuis 2021, l'équipe universitaire ici rassemblée participe à un projet international de recherche comparative financé par le programme ORA (Open Research Area). Intitulé *Shaping 21 st Century AI : Controversies and Closure in Media, Policy, and Research*, le projet implique la participation de quatre équipes, situées dans quatre pays : la France, l'Allemagne, la Grande-Bretagne et le Canada. L'objectif principal est d'examiner comment se construit socialement l'IA, c'est-à-dire à travers quelles focales et selon quels niveaux de problématicité et de controversialité. L'hypothèse de base est la suivante : cette construction n'a pas les allures d'un long fleuve tranquille, mais est plutôt constituée d'épisodes de dissensus et même de controverse – bien que ceux-ci ont tendance à se résorber via différents mécanismes de fermeture (*closure*). Quatre niveaux – les médias, le politique, la recherche et l'engagement du public – structurent la collaboration au sein du projet. Sur une période de trois ans, chaque équipe examine, dans son propre contexte national, comment chacune de ces couches façonne la compréhension sociétale quant aux avancées de l'IA. Aussi, l'attention portée permet à chaque équipe de procéder à une analyse comparative des quatre couches.

Ici, il importe surtout de mieux situer le rapport avec les deux qui le précèdent quant à la représentation de l'IA dans les médias canadiens (Dandurand et coll., 2022) et la manière dont les cercles politiques et l'État se sont saisis de la question (McKelvey et coll., 2024). Dans le rapport *Training the News : Coverage of Canada's AI Hype Cycle (2012-2021)*, Dandurand et coll. analysent les positions, pratiques et processus de fabrication de l'information qui se prêtent à la couverture des nouvelles scientifiques et technologiques au pays. Deux méthodes de recherche principales ont été utilisées : des entretiens semi-structurés avec des journalistes (n=14) et une analyse informatique, utilisant les techniques

de reconnaissance des entités nommées et de modélisation des sujets traités. Elles ont permis de montrer qu'au Canada, les représentations de l'IA dans les médias traditionnels reflètent principalement – et le plus souvent sans grande problématisation – les intérêts des entreprises ainsi que la capacité pour ainsi dire « promise » de l'IA à apporter une future croissance économique. En parallèle, l'analyse des médias canadiens a montré comment la couverture de l'IA suit un cycle d'engouement sous la forme d'un battage *médiatique*, à savoir que : i) les nouvelles technologiques ont tendance à adopter un ton techno-optimiste et proviennent généralement du domaine des affaires et non de ceux de la science ou de la technologie ; ii) il n'existe pas d'écarts significatifs dans la couverture de l'IA entre les salles de rédaction anglaises et françaises ; iii) les gadgets, les voitures autonomes ou autres applications courantes intéressent davantage les journalistes que les nuances sociales ou techniques ; iv) les informaticien.nes sont les expert.es le plus souvent représenté.es dans les médias en matière d'IA ; v) les médias n'examinent que peu, ou pas du tout les modalités du financement de la recherche en IA au Canada ; vi) l'éthique domine le discours public sur l'IA dans les médias traditionnels ; vii) les éditeurs de presse s'appuient de plus en plus sur des technologies d'IA dans leurs fonctions, mais ne discutent pas ou très peu des implications de l'IA pour le journalisme. À terme, ces résultats sont non seulement significatifs en eux-mêmes, mais permettent de mieux comprendre comment les chercheur.es canadien.nes s'inscrivent aujourd'hui dans une dynamique contextuelle particulière ; dite dynamique que leurs travaux façonnent sans doute autant que celle-ci les façonne.

Dans *Northern Lights and Silicon Dreams: AI Governance in Canada (2011-2022)*, ce sont à la fois les différentes manières par lesquelles l'engouement et les mythes de l'IA influencent les cercles politiques et comment ces cercles ont en retour participer, sinon promu, ce même engouement qui sont étudiés (McKelvey, Toupin, Roberge et coll., 2024). Ce sont ainsi quatre études de cas qui sont mises de l'avant : i) L'impact algorithmique du Canada ii) L'approvisionnement en technologies de l'IA par le gouvernement fédéral, iii) Consultations à propos de l'emploi des technologies de reconnaissance faciale, iv) le Canadian Institute for Advanced Research (CIFAR). Chacune de ces études de cas met en lumière un site particulier d'intervention technologique pour l'État et par lequel la technologie en vient à (re)façonner la gestion de l'État en elle-même. Les auteurs remarquent que les interventions gouvernementales sont principalement économiques et visent à notamment à créer un environnement dans lequel les industries pourraient adopter et exploiter toujours davantage d'IA. Ils notent aussi que l'IA est étroitement liée à l'autopromotion du Canada et du Québec en tant que « leader », faisant de l'IA une

technologie nationaliste, bien que « responsable ». Enfin, il s'agit de voir que la gouvernance de l'IA dans le fonctionnement même de l'État est peu coordonnée, laissant la conception, le développement et le déploiement de cette technologie entre les mains d'ingénieur.es et d'informaticien.nes. Comme le montre la construction de la loi sur l'intelligence artificielle et les données (LIAD), les initiatives actuelles en matière de réglementation de l'IA sont plus susceptibles d'accommoder les industries y travaillant que de promouvoir un nouveau format de consultation ou, d'ailleurs, un nouveau type de relation entre les processus démocratiques et technologiques.

1.2 Cadre théorique et méthodologique

1.2.1 Quelques pistes proposées dans la littérature en étude des sciences et des technologies

La présente étude n'est certes pas la première à en appeler d'une conception holistique du phénomène « IA », à savoir d'une prise en compte qui est tout à la fois technologique, sociale, culturelle, économique et politique. À titre d'exemple, les exercices de Bloomfield – « The culture of Artificial Intelligence » (1987) – ou de Woolgar – « Why not a sociology of Machine » (1985) – allaient déjà en ce sens à l'ère des systèmes experts. Aujourd'hui, c'est ce même appel qui se trouve par exemple dans l'idée de proposer une « End-to-end sociology of contemporary AI/ML » (Roberge et Castelle, 2021). Commune ainsi à toutes ces tentatives est la reconnaissance du caractère à la fois dynamique et pluriel de ce qui est en jeu dans la construction et le déploiement de l'IA. Ces enjeux sont à proprement parler « joués » ; non pas qu'ils soient frivoles, mais plutôt qu'ils soient perpétuellement négociés et performés. Cela rejoint la formule chez Pinch et Bijker évoquée ci-haut d'une « flexibilité interprétative » (1984) à même de mouvoir les acteurs impliqués dans le développement technologique, y compris celui de l'IA ici et maintenant. Dans la recherche faite au Canada, les discours forgent les pratiques comme les pratiques forgent les discours ; plus encore, c'est leur enroulement qui importe. En termes de sociologie des sciences ou des études des sciences et technologies (STS) plus classiques, ce type « d'embobinement » est notamment présent dans la conception mertonnienne des *prophéties autoréalisatrices* (Merton, 1948) ou celle, callonienne et latourienne, de *traduction* (Akrich et coll., 2002). (Re)dire, mettre en mots, c'est s'adapter en contexte d'incertitude quant à ce qui se trame et ce qui est en jeu. C'est trouver un langage spécifique capable de rendre compte des excitations et appréhensions ; c'est transformer les justifications en revendications légitimes, les traducteurs en entrepreneurs discursifs, etc. Mais, encore une fois, sans

que les effets ne soient parfaitement contrôlables en sorte qu'une forme d'ambiguïté demeure toujours. Le champ de l'IA est emblématique à ce propos (Hoffman, 2017). Comme mélange de sciences et de technologies de *pointe*, comme intrication de discours et de pratiques, de réalisations et d'imaginaires, c'est jusqu'à sa définition qui est perpétuellement instable. Entre autres choses, une vaste partie du premier chapitre du rapport s'attache à cet enjeu de savoir de quoi parlent les chercheur.es canadien.nes lorsqu'ils parlent d'IA.

Une autre source d'inspiration pour ainsi dire « classique » utilisée dans ce rapport relève des travaux en ethnographie des laboratoires aux débuts des années 1990, ceux de Diana E. Forsythe en particulier. Dans son ethnographie des laboratoires informatiques, elle examine comment les ingénieurs en intelligence artificielle adoptent une position épistémologique et ontologique lorsqu'ils tentent de donner un sens aux programmes cherchant à imiter la prise de décision humaine via des systèmes experts. Elle écrit :

> As scientists construct artifacts and meanings, they draw upon a repertoire of familiar beliefs about the way the world is ordered, some of which are explicit and some of which remain tacit. Among other things, such meaning embodies understandings about what can be a problem and what can be a solution. An important concern […] is to illuminate these beliefs, and to investigate their relationship to scientific practice (1993, p. 450).

L'argumentaire Forsythe est en ce sens pleinement constructiviste : dans le domaine de l'IA, le savoir est interprétation, sélection et traduction. Elle va même plus loin en parlant d'un *éthos* bien particulier appartenant aux ingénieur.es et autres spécialistes des sciences computationnelles (1993 : p.456). *Primo*, existe cette tendance à concevoir les problèmes comme étant pratiques et techniques plutôt que théoriques. Cette forme de positivisme permet alors la mise en avant de certains critères et caractères épurés, sinon simples : l'analogie entre intelligence-machine et cerveau humain par exemple ou l'emphase sur la performance des modèles d'IA comme il en sera abondamment question dans le premier chapitre du rapport. *Secundo*, toujours pour Forsythe, il existe une certaine tendance dans cette culture de laboratoire à réifier la connaissance, c'est-à-dire à penser que quantification et automatisation vont de pair tout en étant suffisants – ce qui, de fait, peut couper court à la nature sociale des mondes qui sont devenus aujourd'hui l'objet de l'IA telles les plateformes numériques, la musique, la bourse, ou comme cela peut couper court à la nature sociale des positions en recherche. L'éthos est bien ainsi un ensemble de

valeurs qu'il s'agit d'actualiser et de réactualiser. Aussi, ces dernières années, sans doute l'article le plus important à ce propos est celui de Birhane et coll. portant sur « The Values Encoded in Machine Learning Research » (2022). Les auteur.es notent que les personnes et les publications les plus influentes

> not only favor the needs of research communities and large firms over broader social needs, but also that they take this favoritism for granted, not acknowledging critiques or alternatives. The favoritism manifests in the choice of projects, the lack of consideration of potential negative impacts, and the prioritization and operationalization of value such as performance, generalization, efficiency, and novelty. These values are operationalized in ways that disfavor societal needs (2022, p. 15).

Qu'en est-il dans un contexte plus spécifiquement canadien, dit contexte qui est marqué par l'historique évoqué ci-haut et différentes déclinaisons notamment en termes de centres/périphéries – la trilogie Mila – Vector – Amii – ou de dynamique linguistique ? Le présent rapport cherche à répondre à ce type de questionnement davantage situé.

Parlant de négociation-traduction, d'ambiguïté et de valeurs, c'est certes un espace de tensions qui se trouve entrouvert. L'hypothèse de base de tout le projet Shaping AI est que cette construction est faite d'épisodes de dissensus, sinon de controverse en sorte que ce soit cette problématicité controversée qui a été recherchée dans le contexte des quatre pays représentés. Qu'en est-il par exemple des diverses réceptions des controverses associées aux *Stochatic Parrots* (Bender et coll., 2021) ou au *Gaydar* (Wang et Kosinski, 2017) et comment font-elles écho au débat plus ancien de la campagne *Against Killer Robots* (Roberge et coll., 2020) ? Qu'en est-il – comme il en sera beaucoup question au chapitre 3 – des différentes positions autour de l'appel du Future of Life Institute pour un moratoire de dix-huit mois sur le développement de l'AI générative ? Ces exemples montrent minimalement une forme d'*agon*, de conflit ou de débat. Ils sont symptomatiques d'une certaine mise en visibilité, en sens et en action de l'IA comme « problème public » (Annany, 2024 ; Bellon et Velkovska, 2023). Tout le problème, ceci étant, est que cette capacité est pour le moins *relative*. Comme le souligne le commentaire de Lucy Suchman, il y a aujourd'hui une forme d'acceptation de ce qu'est l'intelligence artificielle qui relève de son « uncontroversial thingness » (2023). L'objet est instable, mais ses différentes oscillations sont assez subtiles et complexes pour ne pas prendre les allures de grandes oppositions ou de clivages insurmontables. Lorsque de nouveaux enjeux émergent, ils

parviennent à se stabiliser ou du moins à demeurer dans une forme d'ambiguïté ; dite forme qui, encore une fois, est au cœur des différents chapitres du présent rapport.

De manière assez surprenante, les entrevues réalisées ont révélé que plusieurs des personnes interrogées se questionnent et expriment des préoccupations similaires quant à la façon dont l'IA est développée et déployée dans la société canadienne. Cela a entre autres tendance à montrer qu'il n'y a pas tant une dichotomie entre par exemple les spécialités computo-centriques et socio-centriques ; la relation entre ces dernières apparaissant plutôt comme ayant l'allure d'un arc ou d'un continuum. C'est donc encore une fois aux détails des différentes nuances, problématisations et même critiques auxquels le rapport s'attache. Par exemple, il existe un certain consensus quant à la nécessité des spécialistes de toutes les disciplines s'intéressant à l'IA à « travailler ensemble ». Qu'est-il entendu exactement par-là ? L'accord tient-il parce que suffisamment vague ou est-il construit en creux des valeurs invoquées ci-haut par Birhane et coll. ? Est-ce qu'autrement dit, de manière plus prosaïque, « le diable est dans les détails » d'une cooptation-récupération possible de qui dit quoi, comment et pourquoi ? Un des objectifs principaux de la présente étude a ainsi toujours été d'aller à la rencontre d'une vaste gamme de chercheur.es, c'est-à-dire de parcourir le terrain de la recherche en IA au Canada, d'y rencontrer les gens qui l'animent.

1.2.2 Précis de la démarche méthodologique appliquée

Au cours de l'été 2023, dix-neuf entretiens semi-structurés ont été menés pour comprendre les récits qui ont façonné la compréhension scientifique et populaire de l'IA, et ce, à l'intérieur comme à l'extérieur des domaines universitaires. Les critères de sélection étaient assez larges : 89 spécialistes ont été contacté.es, issu.es à la fois des sciences sociales et des sciences, de la technologie, de l'ingénierie et des mathématiques (STEM) travaillant dans une institution canadienne et ayant une expertise contributive en IA ou dans des domaines connexes, à savoir une expertise en réseaux neuronaux, en éthique de l'IA ou dans les composantes discursives des technologies[16].

[16] Dans le cadre du travail comparatif de Shaping AI, nous avions précédemment établi des listes de 700 spécialistes de l'IA au Canada. La première liste était basée sur les universitaires locaux ayant présenté un article lors d'une conférence nationale ou internationale et comprenait 291 noms. La seconde liste comprend des universitaires d'institutions canadiennes qui déclarent travailler sur l'IA et les questions connexes. Elle comporte 418 noms. Nous avons utilisé les deux listes pour sélectionner les participants aux entretiens semi-structurés. Près de 45 % des experts interrogés s'identifient comme des femmes et parlent français, et 15 % appartiennent à des communautés noires, autochtones et de couleur. La plupart d'entre eux sont issus des domaines des STIM (63 %). La répartition géographique des experts interrogés au Canada est également biaisée en faveur du Québec (47 %), suivi de l'Ontario (21 %), de l'Alberta (16 %), de la Colombie-Britannique (11 %) et des Maritimes (5 %). Notre échantillon d'experts interrogés est assez proche des caractéristiques de la population d'experts en IA au Canada.

Les entretiens semi-structurés ont été menés par vidéoconférence et ont duré entre 60 et 180 minutes. Au total, six thèmes ont été abordés au cours des entretiens : i) la position de la personne interrogée dans le domaine ; ii) l'histoire de l'IA ; ii) les défis techniques et scientifiques de la construction de l'IA en laboratoire ; iv) la transition de l'IA des laboratoires à la société ; v) les impacts sociaux de l'IA ; et vi) sa pérennité. Les entretiens ont ensuite été enregistrés, transcrits automatiquement à l'aide du logiciel Descript, codés et analysés de manière collaborative lors d'un atelier[17].

Le cadre méthodologique repose ainsi surtout sur des entretiens semi-structurés. Les idées des chercheur.es ont été recueillies quant à divers éléments de l'IA afin de sonder et d'interroger les différents points de vue, enjeux et récits sur l'IA. Comme indiqué précédemment, un réseau d'acteurs, d'institutions et d'organisations – qui s'efforce de faire de l'IA une ressource clé pour la prospérité économique et sociale du Canada (Colleret et Gingras, 2022) – a façonné la compréhension collective de la technologie. Cet imaginaire est celui du CIFAR par exemple et a ainsi relativement bien trouvé écho dans les cercles politiques – à ISED notamment – de même que dans les grands médias. En cherchant à savoir comment les chercheur.es situent un objet technoscientifique comme l'IA au Canada aujourd'hui, l'objectif était d'examiner de manière critique un site pour ainsi dire en amont. Pour ce faire, place a été faite aux acteur.trices qui construisent et donnent un sens à ces modèles en premier lieu. Comme vu ci-haut, la production de connaissances est située au sens suggéré par Diana E. Forsythe (1993) en sorte que porter l'attention sur les situations et les controverses liées à l'IA telles qu'elles sont perçues donne un aperçu des socialités et des contextes culturels qui sous-tendent ces systèmes sociotechniques.

Comme le montre le présent rapport, une multitude de personnes sont intimement intégrées à une pluralité de systèmes d'IA pour façonner leur conception et leur déploiement (Holton et Boyd, 2019 ; Seaver, 2017 ; 2018 ; Christin, 2020). L'intérêt à sonder le rôle qu'elles jouent dans l'écosystème canadien de l'IA vient de la conviction que ces personnes sont à la fois très proches et pourtant assez éloignées de leur objet de recherche. En effet, toutes les personnes interrogées possèdent une expertise contributive à leur champ de connaissances (Claveau et Prudhomme, 2017 ; Collins et Evans, 2002 ; 2007). Au fil des ans, elles ont acquis une proximité avec l'IA – qu'elle soit épistémologique, ontologique, politique ou un assemblage des trois – qui leur a permis

[17] Un interlocuteur nous a demandé que son enregistrement ne soit pas stocké sur les serveurs d'une société privée. Par conséquent, nous n'avons pas utilisé les capacités d'enregistrement en nuage de Zoom ni la fonction de transcription de *Descript*. L'entretien a été traité localement, en utilisant *Whispers* comme logiciel de transcription.

d'articuler avec des niveaux élevés de granularité et de positionnalité son fonctionnement et l'impact de ces systèmes sur la société. Toutefois, comme l'expliquent lucidement Robert Holton et Ross Boyd, il existe également une distance – ce qu'ils appellent une « distance cognitive » (2019, 181) – entre ce que ces personnes expertes cherchent à générer, ou les résultats qu'elles cherchent à créer et l'opérationnalisation des algorithmes dans le monde. Les chercheur.es ont certainement un pouvoir d'action dans la construction des systèmes d'IA, mais ce pouvoir dépend de la complexité des systèmes et du contexte culturel dans lequel ils sont censés fonctionner. C'est en partie l'objectif de ce projet de recherche : il donne aux personnes expertes l'occasion de réfléchir à leur propre objet de recherche et au domaine qu'elles contribuent à créer, en partageant leurs analyses sur la manière dont ces systèmes d'IA laissent des traces sociomatérielles dans le monde.

Dans les chapitres suivants, il s'agira de voir comment les personnes interrogées reconnaissent d'une manière ou d'une autre les dynamiques de pouvoir qui structurent la conception et le déploiement de l'IA. La plupart des personnes interrogées se sont avérées assez critiques à l'égard de la relation entre le complexe industriel transnational de l'IA et les départements universitaires d'informatique canadiens. D'autres n'hésitent pas à apporter des nuances à ce qui pourrait être perçu comme un débat public stérile et une promotion de l'« IA responsable ». Beaucoup remettent en question la probité des « risques existentiels » qui semblent structurer d'une certaine manière les débats actuels sur l'avenir de l'IA, soulignant au passage le besoin pressant de faire quelque chose au sujet des cas bien documentés de préjudices posés par l'IA qui s'avèrent de plus en plus être une caractéristique commune à plusieurs sous-types de la technologie.

Ces points de vue sur l'IA peuvent être biaisés. Comme indiqué ci-haut, l'équipe de recherche a envoyé des invitations à 89 personnes expertes en IA dans tout le Canada, y compris au grand nombre d'informaticien.es financé.es par le CIFAR et affilié.es à l'AMII, au MILA ou à l'Institut Vector. En termes relationnels, un nombre étonnamment faible de réponses de la part des membres du MILA et de l'Institut Vecteur a été reçu. Bien qu'il soit difficile de le savoir, il est soupçonné que les personnes expertes qui ont donné une à trois heures de leur temps pour parler avaient préalablement des inquiétudes ou des insatisfactions quant à l'état actuel de la recherche sur l'IA au Canada et cherchaient de ce fait une occasion de les partager. D'autres personnes qui bénéficient de l'infrastructure actuelle de financement de la recherche en IA ne sont peut-être pas aussi enthousiastes pour parler de questions qui structurent cet écosystème particulier.

1.3. Organisation des chapitres

Le premier chapitre du rapport se penche sur les débats épistémiques concernant la définition et la constitution des modèles d'IA ainsi que sur la relation privilégiée entretenue entre cesdits modèles et les manières d'en parler dans la communauté scientifique – qu'il s'agisse des spécialistes provenant du champ technoscientifique de l'apprentissage automatique ou de celles et ceux provenant des sciences sociales. Il est question d'abord d'analyser les déclarations des différents intervenant.es rencontré.es concernant les problèmes générés par l'*indéfinition* du champ de l'IA, le caractère instrumental de l'interprétation automatique du langage, les caractéristiques structurelles des modèles ainsi que les controverses reliées à l'interprétabilité de ceux-ci.

Les témoignages des personnes rencontrées au cours de ce projet de recherche permettent de constater que la définition même de l'IA ne fait pas consensus parmi ceux-ci. Cette indéfinition fondamentale caractérisant le champ technoscientifique renvoie elle-même à une sous-série de débats caractérisant cette discipline. Les progrès fulgurants récemment effectués dans le domaine spécialisé du traitement naturel du langage font ressortir qu'il n'existe pas de consensus au sein de la communauté scientifique quant à la valeur effective de la notion de compréhension. De plus, bien que certains acteurs de la communauté scientifique avancent que la croissance des modèles d'IA est leur principal facteur de réussite quant à la résolution de problèmes, plusieurs spécialistes rencontré.es lors des entretiens critiquent ce type de discours, croyant au contraire que la taille de ces modèles ainsi que les procédés employés pour les constituer nuiront à long terme au développement de leur discipline. L'opacité des modèles lorsque vient le moment de produire des explications à propos de leur fonctionnement est non seulement un facteur qui nuit à leur déploiement dans des contextes de prise de décision critique, mais cause également des problèmes sur le strict plan épistémique. Produisant une connaissance dépourvue des explications qui devraient normalement l'accompagner, ces modèles alimentent un débat fondamental concernant leur interprétabilité. La nécessité de cette dernière faisant elle-même l'objet de contestations au sein de la communauté scientifique, sa définition objective ainsi que les méthodes permettant de l'atteindre ne font pas non plus consensus. Si la définition du champ de l'IA, la constitution de ses modèles et leur interprétation provoquent autant de dissensions, il est conséquent que ces désaccords influencent les échanges entre les experts et créent des répercussions sur les paroles et les écrits circulant dans la communauté scientifique.

La suite de ce premier chapitre du rapport se penche justement sur le dialogue réunissant les scientifiques autour de la thématique de l'IA. Cela est découpé en trois sous-ensembles décrivant ce dialogue par couches successives, à la manière de trois cercles concentriques. Le premier des sous-ensembles, jouant le rôle de noyau, est constitué par le langage employé pour décrire l'IA par les membres de la communauté scientifique. Selon les spécialistes, l'introduction rapide de nouveaux concepts couplée à une transmission défaillante de l'histoire de la discipline provoquerait des enjeux communicationnels s'incarnant notamment à travers de nombreux glissements terminologiques. Le deuxième des sous-ensembles englobe ce premier noyau et regroupe les experts issus du champ technoscientifique de l'IA, c'est-à-dire les chercheurs évoluant dans des domaines spécialisés *connexes* ou adjacents à l'IA tels que l'informatique. Plusieurs personnes rencontrées lors du projet de recherche décrivent une situation plutôt hétérogène où différents groupes proposant des visions de l'IA et de son devenir s'opposent. Les enjeux proprement langagiers du noyau décrit précédemment contribuent à conférer un caractère discordant à ce dialogue intradisciplinaire. Le dernier des sous-ensembles est constitué par les échanges entre ce premier groupe de scientifiques et les spécialistes des sciences sociales alors que ces dernier.ères constatent une déconnexion assez marquée entre les groupes disciplinaires. Malgré ces difficultés, la majorité des expert.es rencontré.es souhaitent développer une collaboration interdisciplinaire afin de surmonter les embûches provoquées par le déploiement de l'IA. Encore une fois, le caractère indéfini du champ de l'IA ainsi que les conflits d'interprétation qui en résultent est ce qui pourrait nuire aux diverses entreprises de traduction menées par les experts.

Le propre de ce premier chapitre est alors d'explorer la cooccurrence de problèmes affectant la signification et les signifiants de l'IA existante tant au niveau de la constitution des systèmes technologiques, que dans leur traitement du langage pour éventuellement se répercuter dans les échanges entre les acteurs du monde scientifique. Sans faire appel à une causalité directe, la proximité thématique et les liens sémantiques unissant ces différents éléments suggèrent une forme d'influence croisée qui nuit à la stabilisation des discours entourant l'IA.

Quant à lui, le second chapitre du rapport traite principalement des matériaux nécessaires à la recherche académique en intelligence artificielle. Plus précisément, il s'agit d'examiner comment l'engouement actuel pour les modèles dits « larges » – ceux-là mêmes qui sont généralement conçus en contexte industriel – agit à titre de facteur structurant de la

recherche académique. Cet engouement et sa force structurante s'offrent comme un « régime de l'*upscaling* » et s'expliquent à la fois par la couverture médiatique de l'IA qui publicise généralement les modèles industriels de type ChatGPT ainsi que par la présence en sol canadien de ces mêmes industries – Google, Meta, etc. – agissant souvent à titre de partenaires de la recherche universitaire.

De fait, les entrevues révèlent que ce contexte exerce une certaine pression sur les chercheur.es académiques les incitant à emboîter le pas des industries en travaillant sur des modèles de taille plus larges, ou en accroissant la taille des modèles sur lesquels ils travaillent déjà. Plus encore, l'organisation même de l'écosystème de recherche subventionné par le gouvernement canadien contribue à accroître cette même pression chez les chercheur.es dont les instituts sont situés à *l'extérieur* du circuit financé par le CIFAR. Les entrevues ont aussi révélé que l'accroissement des tailles des modèles nécessite parallèlement l'accroissement quantitatif de trois ressources : le financement, les données et la puissance de calcul. Si l'accès à ces ressources en quantité suffisante semble d'une relative difficulté pour les chercheur.es opérant à l'intérieur du circuit supporté par le CIFAR, il est en effet encore plus difficile pour celles et ceux opérant à l'extérieur de ce même circuit qui doivent dès lors multiplier les stratégies pour continuer de travailler sur leurs projets. Entre autres, ce chapitre est l'occasion de s'interroger sur les conséquences de cet accroissement des ressources sur les pratiques en laboratoires, mais aussi, sur l'ensemble de la société et ce que cela pourrait, en outre, venir signifier en termes de *souveraineté numérique* pour le Canada.

Enfin, le troisième et dernier chapitre du rapport rassemble les réponses des personnes rencontrées en ce qui a trait à l'impact des systèmes d'IA une fois déployés au-delà des laboratoires. Il présente ainsi trois thèmes principaux : les discours promotionnels de l'IA, ses risques sociaux et politiques de même que la question de la gouvernance de ce type de technologie dans les sociétés démocratiques. En ce qui concerne les discours promotionnels touchant aux prouesses l'IA, les expert.es interrogé.es partagent largement le point de vue selon lequel le battage médiatique actuel est devenu tout aussi répandu que problématique. De nombreuses personnes partent du même constat : le public n'est pas informé de la réalité et des capacités des systèmes d'intelligence artificielle, ce qui les rend vulnérables au matraquage publicitaire et à différentes formes d'utilisation qui pourraient être jugées abusives. Plusieurs personnes rencontrées ont ainsi déploré le fait que l'IA soit considérée comme une solution « quasi magique » pour tout un éventail de choses les plus variées, y compris par les journalistes qui, selon encore

une fois les personnes rencontrées, pourraient ou devraient être davantage critiques. Une préoccupation majeure soulignée est en l'occurrence la façon dont ce battage médiatique a également un impact sur le développement scientifique du domaine de l'IA lui-même. C'est que l'évolution rapide des vocabulaires, des mots à la mode et des imaginaires utilisés pour commercialiser et discuter publiquement de l'IA façonne en retour la manière dont les étudiant.es et les ingénieur.es de l'IA comprennent ces systèmes et leurs capacités. Plus problématique encore, la course à la publication d'articles de recherche de pointe et au dépassement des critères de référence pousse de nombreuses personnes œuvrant dans le domaine à faire de la science dite « bâclée » et à ajuster les modèles de manière à suivre ce battage médiatique plutôt qu'autre chose.

Parlant des risques sociaux et politiques de l'IA, il existe une controverse considérable entre les discours fortement médiatisés et les points de vue des personnes interrogées. Par exemple, si l'automatisation des emplois et l'utilisation de l'IA pour développer des armes militaires suscitent des inquiétudes, la plupart des personnes rencontrées estiment que le discours sur les « risques existentiels » promus par les Parrains de l'IA tels que Joshua Bengio et Geoffrey Hinton, entre autres, est exagéré et détourne l'attention de questions plus immédiates. En fait, l'un des principaux sujets de préoccupation relevés concerne les risques que l'IA générative fait peser sur l'écosystème de l'information. Les personnes rencontrées s'inquiètent non seulement de la désinformation intentionnelle générée et diffusée par de « mauvais acteurs », mais aussi des risques de désinformation involontaire ou accidentelle créée par des *robots* d'IA générative peu performants. Dans l'ensemble, les personnes rencontrées sont beaucoup plus préoccupées par les hallucinations et les modèles qui ne fonctionnent pas aussi bien que souhaité que par les menaces existentielles lointaines que représenteraient des systèmes d'IA « trop intelligents ».

Ce troisième et dernier chapitre aborde par ailleurs la question de la gouvernance de l'IA. Sont présentés les points de vue des spécialistes quant aux personnes qui devraient être entendues et qui devraient pour ainsi dire avoir un siège à la table pour façonner le développement de l'IA, depuis la conception jusqu'au déploiement et à la réglementation. Les spécialistes ont massivement dénoncé le manque de compétences et de connaissances en matière d'IA parmi les agences gouvernementales, affirmant que des équipes d'informaticien.nes et de spécialistes des sciences sociales devraient être impliquées dans les efforts de réglementation à chaque étape. Plusieurs ont même défendu avec passion la nécessité d'inclure un large éventail d'experts disciplinaires –

spécialistes des sciences sociales, des sciences humaines, de l'histoire, etc. – dans les équipes de recherche sur l'IA afin de contribuer aux efforts de développement dès leur début. Enfin, la question de savoir si le public doit être associé à ces conversations et à ces efforts se montre être un sujet de controverse et de désaccord. Alors que certaines personnes interrogées ne voyaient pas l'intérêt d'inclure des points de vue non spécialisés, beaucoup d'autres estiment qu'un groupe diversifié de voix provenant de l'ensemble de la société devait être inclus et consulté afin de garantir le développement responsable des systèmes d'IA et de veiller à ce que les personnes qui sont déjà les plus marginalisées dans la société ne subissent pas d'autres conséquences négatives du déploiement de l'IA.

2. PREMIER CHAPITRE — Débats épistémiques et les controverses de l'interprétabilité

L'IA est fréquemment présentée comme une technologie généraliste pouvant être déployée dans des contextes variés, ses capacités prédictives s'appliquant à des problèmes qui pourraient d'emblée paraître éloignés des sciences informatiques (Dyer-Witheford et coll., 2019). Elle investit les domaines de la santé et du droit, comme en ont témoignés plusieurs des personnes rencontrées par l'équipe canadienne de *Shaping AI*. Cette propension à investir des milieux hétérogènes conduit à conceptualiser l'IA comme un assemblage de solutions technologiques n'ayant parfois, sinon souvent qu'un couplage flou avec les problèmes à résoudre. « What is the problem for which these technologies are a solution? According to whom? », se demande l'anthropologue Lucy Suchman (2023) alors qu'elle investigue l'indéfinition caractérisant l'IA. Cette particularité du champ technoscientifique peut s'expliquer en ces termes : « AI currently enjoys a profound as well as multifaceted hype that might be rooted in the sort of ambiguity that comes with an uncertain and contingent future. Hype, ambiguity, and efficiency go hand in hand » (Roberge et coll., 2020).

Afin de comprendre l'origine de cette indéfinition caractérisant l'IA, il importe de s'intéresser aux multiples discours véhiculés par les différents groupes de scientifiques associés à son développement, qu'il s'agisse de ceux situés du cœur même de la discipline technoscientifique ou davantage à une certaine périphérie correspondant – de manière plus ou moins surprenante – à son point d'entrée dans la société. Ces groupes participent d'une forme commune ; ils coopèrent à travers un certain *modus operandi* et une certaine production discursive sans que ne se fasse sentir la nécessité d'un consensus à propos du problème scientifique ciblé ou même d'une coordination explicite. Autrement dit, tout se passe comme s'ils produisaient ensemble des traductions (Callon, 1984 ; Latour, 2005) concernant les paramètres d'un problème scientifique en lien avec les caractéristiques du monde auquel ils appartiennent. Ainsi, nombre d'expert.es provenant des disciplines STEM, même différenciés par leur appartenance à un sous-champ spécialisé de l'IA,

qu'il s'agisse de traitement du langage, de vision par ordinateur, d'analyse des données, ou de toute autre déclinaison récente de cette branche de l'informatique, entrent en relation avec des scientifiques provenant des sciences sociales dont le principal objet de recherche est le déploiement de l'IA dans la société, la culture, l'économie, etc. Ces scientifiques différencié.es par leur propre sous-champ disciplinaire (anthropologie, sociologie, philosophie, droit, santé publique, etc.) sont parties liées à une discussion plus large où figurent également des non-expert.es, soit des individus qui entrent, volontairement ou non, en contact avec une IA déployée à travers une pléthore d'objets techniques faisant maintenant partie de l'environnement.

Cette portion du rapport de recherche est consacrée aux enjeux épistémiques soulevés par l'IA et les effets que ceux-ci produisent sur une certaine conversation sociale. L'indéfinition de l'IA ainsi que sa relation historique paradoxale avec l'interprétation du langage sont des pistes qui permettent d'expliquer une partie des résultats obtenus par l'équipe canadienne de Shaping AI. Une première section explore les débats épistémiques associés à l'IA tels qu'ils sont perçus par la communauté scientifique canadienne. Une deuxième section présente la traduction effectuée par cette même communauté quant aux débats décrits précédemment. Cristallisés dans les controverses associées à l'interprétabilité, ils font l'objet de tentatives de résolution à travers l'implantation d'un cadre technique qui permettrait de dissiper en partie l'indéfinition de l'IA. Le chapitre est alors consacré à trois moments distincts. D'abord, ce sont les enjeux concernant la terminologie employée par les experts qui sont présentés. Ensuite, il est question d'étudier les qualités du dialogue intradisciplinaire au sein du champ technoscientifique de l'IA. Enfin, cette même analyse est élargie par l'inclusion de la question des échanges interdisciplinaires, notamment ceux impliquant les représentants des sciences sociales.

2.1 L'indéfinition de l'IA, carburant des controverses

Au tout début du canevas d'entrevue utilisé par l'équipe de recherche se trouvent des questions permettant de sonder rapidement les personnes rencontrées à propos des principaux enjeux ou controverses associés à l'IA[18]. Plutôt que de suggérer une controverse spécifique telle que le *Gaydar* ou les *Killer Robots*[19], ces questions sont ouvertes et

[18] Voir la partie méthodologique de l'introduction pour davantage de détails.

[19] Ces deux controverses associées à l'IA concernent des cas relayés par la presse généraliste et certaines publications spécialisées et font respectivement référence à la détection de l'orientation sexuelle par vision informatique et à l'automatisation de drones aux capacités léthales. Ces controverses ont été spécifiquement identifiées dans le contexte du sondage envoyé massivement par les différentes équipes nationales de l'initiative Shaping AI qui fera l'objet d'une publication scientifique différenciée. Voir aussi Marres et coll. 2024.

laissent les personnes s'exprimer spontanément. Plusieurs d'entre elles, qu'elles soient associées aux sciences informatiques ou aux sciences sociales, ont d'emblée identifié l'indéfinition de l'IA comme un des principaux problèmes affectant le domaine. Certains individus interviewés refusent littéralement à l'IA le statut de discipline scientifique ; ce type de définition, sans être majoritaire, ayant pour effet de repositionner l'IA en tant que sous-domaine de l'ingénierie :

> People talk about building an AI. I think that's ridiculous. That's not science. You're just building, I don't know, a widget, right? […] It's not an academic discipline in the same sense. It's an engineering problem and, perhaps, even an applied engineering problem (exm9d1).

Considérée comme un assemblage plus ou moins structuré de technologies diverses, certaines personnes avancent que de parler d'IA, « ça ne renvoie à rien de précis comme étiquette » (glv2br). Plusieurs personnes ont évoqué une IA composée d'une multitude de programmes informatiques, d'appareils, de bases de données, voire de « bidules » (glv2br). Il devient dès lors difficile de tracer les contours de ces agencements technologiques et d'en fixer l'identité.

Cette indéfinition de l'IA apparait conférer un avantage à la fois discursif et stratégique aux chercheur.es dans la mesure où son caractère flou permet de lui accoler des significations variées, à savoir sans être gênée par une identité trop forte qui aurait établi avec précision ce qu'elle est. Simultanément, le terme « IA » est capable de mobiliser efficacement un ensemble de croyances, de peurs et d'espoirs plus ou moins irrationnel. « C'est très accrocheur, ça frappe l'esprit » déclare une des personnes rencontrées (glv2br). Cette dynamique particulière alliant la stimulation de l'imaginaire sociotechnique à un canevas destiné à recevoir un ensemble d'idées projetées caractérise de ce fait plutôt bien le développement de l'IA. Une seconde personne interviewée décrit en ces termes le magnétisme de l'IA et son impact sur le financement des activités scientifiques :

> Le terme intelligence artificielle est un terme purement médiatique qui, d'ailleurs, est génial. […] C'est vraiment génial d'avoir inventé cette formulation parce que ça suscite l'intérêt non seulement du public, mais du public fortuné et des organisations qui ont les moyens de subventionner (gk3nxq).

L'absence de définition stricte concernant l'IA permet le foisonnement de récits aux contours plus ou moins mystifiants qui, à leur tour, alimentent son propre développement et ainsi de suite. Cet effet d'entraînement ou de boucle récursive contribue ainsi à maintenir le caractère indéfini de l'IA.

Un contexte scientifique où dominent des experts, comme celui du développement de l'IA pourrait, en théorie, être plus favorable à l'émergence ou au maintien d'une identité plus définie, cohérente et consensuelle de son objet. Or, cette hypothèse a été invalidée par les propos de plusieurs des personnes rencontrées qui décrivent plutôt certaines tensions au sein du champ technoscientifique quant à la définition de l'IA. Une de ces personnes décrit ainsi cette discussion :

> On ne sait pas trop c'est quoi l'intelligence artificielle. Même les gens qui travaillent dans le domaine ont différentes définitions. C'est difficile à ce moment-là de saisir l'objet, précisément parce qu'il n'y a personne qui s'entend exactement sur c'est quoi (j6b3dt).

Cette indéfinition de l'IA, maintes fois relevée lors des entrevues, peut être déclinée sous la forme d'une gradation. Si certains évoquent l'absence de consensus afin d'expliquer les difficultés rencontrées dans l'établissement d'un dialogue intradisciplinaire au sein des sciences computationnelles mêmes, l'inconfort d'autres va jusqu'à même nier l'existence d'un ensemble de connaissances partagées : « There is no mainstream understanding of AI […] within computing » (3jpw66). Pour cette personne, l'absence d'un socle commun de connaissances combiné à l'emploi du terme « IA » en tant que raccourci conduit à une perte de sens qui minerait les efforts des scientifiques :

> I regard the use of AI as a useless waste of time. Because it just, it confuses what's going on. You can't tell what people are talking about because they use this as, I think, what they imagine as a mental shortcut. It's a shortcut to so many different things that it ends up being meaningless (3jpw66).

Ce déficit de sens permet ainsi au moins en partie de comprendre les rapports structurant ce milieu. Entre autres, il convient de s'intéresser aux propos véhiculés par certaines voix discordantes qui témoignent chacune à leur manière d'un problème global touchant la communauté des spécialistes en IA.

2.2 Principaux facteurs contribuant à l'indéfinition de l'IA

La production de données liées aux interactions et à la circulation des usagers, qu'elles soient localisées sur Internet ou pistées à l'aide de téléphones intelligents a nourri une vaste industrie dédiée à leur traitement et leur valorisation durant la première décennie des années 2000. Vers la fin de l'année 2015, l'intérêt suscité par les capacités prédictives associées aux données a poussé les industriels de ce secteur technologique à redéfinir leurs activités et à les présenter sous un nouveau vocable (Elish & boyd, 2017). Passant des mégadonnées à l'IA, l'industrie a en quelque sorte renouvelé sa marque de commerce. Aussi, ce glissement dans les termes utilisés fut simultanément accompagné d'une transformation des objectifs poursuivis par les chercheur.es agissant à la fois dans le monde universitaire et en tant qu'ingénieur.es auprès des plateformes telles que Meta ou Google. Plutôt que de rester ancré dans le programme scientifique traditionnel de l'IA décrit lors de la conférence de Dartmouth évoqué en introduction de ce rapport (Cardon et coll., 2018 ; Mendon-Plasek, 2021), ceux-ci allaient désormais poursuivre d'autres objectifs :

> 'Meaning' is beside the point; the algorithm 'knows' in the sense that it can correlate certain relevant variables accurately. It does not matter if a system thinks like a human–as long as it appears to be as knowledgeable as a human » (Elish & boyd, 2017, pp. 7-8).

Cette double transformation affectant autant la terminologie désignant l'IA que les objectifs poursuivis par le champ de la recherche et les industries contribuent à expliquer en grande partie son indéfinition actuelle. Les propos des personnes interviewées concernant les tensions opposant cognition et compréhension exemplifient avec éloquence la déviation effectuée à partir du programme scientifique historique de la discipline.

La quasi-totalité des participants a rejeté toute forme d'influence exercée par des approches biomimétiques s'inspirant du fonctionnement du cerveau et/ou du développement humain. Comme l'explique l'une des personnes interviewées, il s'agirait plutôt de déployer des systèmes performants dans l'exécution de tâches pour lesquelles l'humain ne l'est pas : « artificial intelligence can do things better than humans can–like derive insights from large amounts of data. […] That is not about doing or imitating what humans do » (exm9d1). Une autre personne abonde dans un sens similaire :

> What is intelligence? […] I refer to this mostly as problem solving and so I'm much more of that camp of saying: "I don't really care whether this is really how a human does it or how an animal does it, or some other kind of cognition does it. I don't care. I only care 'does it solve the problem?'" And there's lots of different ways to try to solve the problem, but *as long as it solves the problem, that's all that I care about. […] That perspective is very common in […] industrial research and artificial intelligence right now. They don't actually care to what extent this tells us anything about human intelligence* (a9a3ir).

Cette perspective sur ce qu'est l'IA place cette dernière dans une certaine instrumentalité où la conformité de sorties en fonction d'objectifs préétablis serait la principale qualité recherchée. Bien que ces systèmes soient capables de générer des contenus symboliques complexes, la question de la compréhension est ainsi évitée, voire décrédibilisée par certain.es chercheur.es :

> I'm using this as a tool to predict the outputs given for these inputs because it's useful. […] Within […] NLP, for example, people often have the same sort of questions, like "Does chat GPT really understand language?" What does it mean to understand? […] *The majority of researchers in the field don't really care*. […] The majority of researchers do not even think about these questions. […] I'm sort of a practical user of the technology. […] I'm not using this answer for any philosophical questions (p9efse).

Ce refus d'aborder frontalement la question de la compréhension et de la cognition humaine montre un certain caractère instrumental, voire positiviste, du déploiement de l'IA et peut sembler laisser de côté les problèmes liés à la sémantique profonde dans l'interprétation automatisée des textes.

Certaines des personnes rencontrées ont souligné que les récents progrès dans le traitement du langage naturel ne pourraient être qu'un phénomène de courte durée. Ces avancées produites par une approche *ascendante* le plus souvent empirique seraient peut-être les derniers fruits produits par cette branche de la recherche en IA :

> Lots of the things that we're doing with deep learning really don't change much, haven't told us much. We haven't learned much from them. They are purely an

> empirical tool and maybe one that gets stale very quickly. We don't really know much about lifespans of these models yet (3jpw66).

Une telle déclaration jette un éclairage différent sur la « fin de la théorie » prophétisée par Chris Anderson (Wired, 2008)[20]. En l'occurrence, l'IA connexionniste d'aujourd'hui ne fait que transposer – plutôt que de « régler » – les enjeux liés à un effort intellectuel auparavant théorique dans un contexte empirique où un modèle probabiliste constitué par le traitement massif de données remplace le précédent. Plusieurs personnes interviewées évoquent même cette transformation comme une perte significative quant à la compréhension des phénomènes constituant le monde. Alors que les corrélations unissant les entrées aux sorties des modèles de l'IA supplantent le précédent régime de causalités, ce type de refus de la théorie prend alors les allures d'un refus du sens et de la signification pouvant être projeté sur ces mêmes phénomènes.

Les linguistes computationnels Emily Bender et Alexander Koller se demandent « are we climbing the right hill? », « not just the hill on whose slope we currently sit », alors qu'ils tentent d'évaluer le progrès réel effectué grâce à l'IA dans leur discipline (2020, p. 5191). Selon ces chercheurs, « maintaining clarity around big picture notions such as meaning and understanding in task design and reporting of experimental results, » est d'une importance capitale afin de s'assurer que les gains incrémentaux réalisés lors de ces expériences guident la recherche dans une voie qui se révélera prometteuse à long terme (2020, p.5185). Cette métaphore de la colline est reprise par l'une des personnes interviewées rencontrées lors des entretiens menés dans le cadre du projet *Shaping AI*. Cette personne explique que la communauté avait rapidement progressé dans l'ascension d'une telle colline, mais que le portrait, une fois au sommet, s'avérait légèrement décevant :

> We were talking about the 80:20 problem. We've put in 20% of the effort and we got 80% of the results. But now, we realized the problem is pretty complex. We cannot get to 100%. And, yeah, this was 2013. And I was saying, well, maybe we finally reached the point where we're going to incorporate context (exm9d1).

[20] Cette thèse a été maintes fois contredites depuis. Comme l'explique Radford et Joseph à propos de l'emploi de techniques provenant de l'apprentissage machine en sciences sociales : « at each step of the machine learning pipeline, problems arise which cannot be solved using a technical solution alone. Instead, we explain how social theory helps us solve problems that arise throughout the process of building and evaluating machine learning models for social data » (Radford et Joseph, 2020, p.2).

Bien que les progrès réalisés récemment dans le domaine du traitement du langage naturel soient stupéfiants, les personnes interviewées identifient une forme de blocage au niveau de la sémantique profonde et de la compréhension. Le *last mile* au terme de cette course dans le traitement du langage apparait difficile à franchir, comme l'affirme une autre personne interviewée :

> Une chose est certaine. […] On n'est pas à l'inférence logique, on n'est pas à la compréhension. On fait de la génération de texte comme on ne l'a jamais fait encore dans le monde […] En tout cas, il y a une chose qui n'a absolument pas fait de progrès notoire, c'est la compréhension. Ces systèmes n'ont pas de compréhension (gk3xq).

Ces propos semblent plaider en faveur de la critique fondamentale formulée par Bender et Koller à propos des récents développements de la linguistique computationnelle. Selon eux, un modèle strictement entraîné sur la forme linguistique n'aura jamais accès à la compréhension textuelle, dans la mesure où les liens unissant le contenu formel à une intentionnalité communicationnelle et à un sens commun partagé (2020). Le bond de géant effectué récemment a peut-être ainsi conduit la discipline sur le sommet de la mauvaise colline. Le caractère parfois, sinon souvent instrumental de la recherche en IA – en guidant son action principalement sur des gains incrémentaux empiriques au détriment d'une attention qui considérerait le phénomène de la compréhension tel qu'il est expérimenté par les êtres humains – pourrait entre autres expliquer cette bifurcation.

Ce problème rencontré dans le traitement du langage à propos de la compréhension du sens profond des textes n'est pas seulement un épiphénomène ; il s'agit d'un enjeu crucial ayant un pouvoir de définition sur la discipline scientifique elle-même. Cette centralité du rapport entre IA, langage et intelligence humaine a ainsi été relevée par une des personnes interviewées :

> it's […] interesting that a lot of the current definition of AI starts with language. We are really impressed by chat GPT because it can do language in a way that sounds human. […] It can generate language that is as good as what a human would say, and therefore it sounds intelligent (exm9d1).

La personne rapporte ici un enchaînement où le rôle du langage dans la définition de l'IA et la capacité des grands modèles de langage (LLMs) à générer des contenus textuels qui

semblent humains résultent en fait d'une projection de qualités humaines sur un système d'IA, ce qui fait en sorte que certains y perçoivent une forme d'intelligence. Le réseau de relations décrit permet ainsi un glissement affectant la notion de compréhension et d'intelligence à propos de systèmes automatisés qui sont justement spécialisés dans le traitement du langage. Comme le soulignaient avec justesse Bender et Koller, l'emploi d'une logique et d'une terminologie parfois, sinon souvent imprécises à propos des LLMs est en partie causé par notre compréhension défaillante des représentations construites par ces modèles à propos du langage (2020).

2.3 Débats épistémiques situés en amont du déploiement des systèmes d'IA

Lors des entretiens, plusieurs individus rencontrés ont souligné qu'une part importante des problèmes causés par l'emploi de l'IA se situe au sein même des laboratoires et, de ce fait, en amont du déploiement des modèles plus largement dans la société. Une conceptualisation initiale limitée des problèmes à traiter ainsi que l'absence de justifications théoriques claires quant au choix d'une méthode induisant un modèle plutôt qu'un autre conduisent une des personnes rencontrées à affirmer que la recherche en IA souffre d'un déficit de nature stratégique :

> One of the […] problems is that people have no idea of strategy. Suppose I have a problem and I want to build a model to solve this problem. I cannot tell in advance what strategy is going to give me the best result, not even in the vaguest possible sense. » (3jpw66)

Ces lacunes en termes de conceptualisation initiale et de théorisation ont également un impact sur la qualité et l'interprétabilité des résultats produits. Cette même personne interviewée conteste d'ailleurs la valeur scientifique même de la recherche en IA déclarant : « that's not really a science. It's barely engineering » (3jpw66). Le traitement des mégadonnées par des systèmes d'IA apparait avoir affecté non seulement les méthodes scientifiques associées au sous-domaine de l'informatique, mais avoir profondément transformé le champ technoscientifique dans son ensemble en biaisant le lien privilégié entretenu entre la production du savoir et une systématisation des causalités pouvant expliquer un phénomène. À ce propos, se remémorant une présentation du MILA auquel il a assisté, un second interviewé commente : « T'as rien expliqué, t'as le résultat là, maintenant, faut le comprendre. » […] Tu as réussi à me dégager les patterns qu'il faut

maintenant que j'étudie pour que j'arrive à comprendre. Je trouve qu'il y a un raccourci mental [...] qui n'a pas d'allure » (x04alc).

Outre ces lacunes concernant la conceptualisation initiale et la théorisation des initiatives de recherche impliquant l'IA, certaines personnes ont soulevé que l'opacité des modèles et l'absence d'une documentation appropriée les rendaient peu fiables et limitaient l'action des chercheur.es. Les efforts communicationnels déployés par les groupes industriels de l'IA à propos de leurs modèles relèveraient parfois plutôt de la sphère promotionnelle que de la communication scientifique a proprement dit selon une des personnes participantes :

> Quand on voit par exemple la note technique de GPT-4 et que c'est plus [...] une brochure commerciale qu'un article scientifique [...], parce qu'on n'a rien sur l'architecture interne, ça pose des questions sur le plan de comment on va gérer ça sur le long terme en recherche en intelligence artificielle (zfr5j8).

L'absence de détails concernant l'architecture de certains modèles nuirait à l'interprétation des résultats qu'ils produisent, selon d'autres. La guerre commerciale que se livre les joueurs industriels de l'IA contribue non seulement à orienter les communications hors du champ technique et scientifique, mais aussi à une discrétion permettant de préserver un avantage sur les concurrents. Une des personnes rencontrées a fourni un exemple de ce basculement des publications industrielles dans le pur registre de la mise en marché tiré de la guerre commerciale que se livrent les industriels de l'IA afin de développer des outils destinés à la production vidéo :

> All the big industry groups are trying to crack video generation and all they're doing is they'll release these [...] short videos on social media. Like, here's all the best examples that we could find of how our model does. And nobody's willing to release it because it's basically an arms race. [...] That's bad (a9a3ir).

Comme l'avance la personne, plusieurs compagnies impliquées dans le déploiement de l'IA à grande échelle refusent d'offrir leurs modèles en téléchargement (et de les décrire précisément) pour plutôt commercialiser un ensemble de services sous forme d'API. L'approche *AI as a service*, autrement dit, contribue à opacifier le fonctionnement des modèles à travers leur fermeture et le déploiement de campagnes médiatiques orientées en fonction de leurs besoins promotionnels plutôt que ceux de la communauté

scientifique. La récupération et l'amplification de ces discours corporatifs par les médias traditionnels (Dandurand et coll., 2024), notamment à travers un traitement principalement effectué sous un angle économique, contribuent à ainsi gonfler le contenu spéculatif, sinon hyperbolique de l'IA plutôt que de favoriser une certaine « mise en lumière » scientifique.

Cette difficulté accrue concernant l'accès à une documentation technique transparente et, de ce fait, au *modus operandi* des modèles d'IA s'ajoute aux problèmes créés par l'expansion des jeux de données et la taille sans cesse grandissante de ces mêmes modèles. Sur un plan épistémologique, en effet, certain.es interviewé.es remettent en doute la validité de ces modèles dans le cadre d'une démarche scientifique, car le nombre de leurs paramètres s'approche de la quantité de données brutes utilisées lors de leur entraînement : « I have philosophical doubts about whether you're doing anything but purely empirical fit when you build models based on that kind of data » (3jpw66). Cette personne expose ainsi assez crûment ce qui peut être l'absence d'apprentissage, de production de connaissances au sein de tels systèmes :

> When the number of those parameters starts to approximate the number of degrees of freedom in the system that you're modeling, then it's really hard to say that you're actually learning some abstraction of the system and not just a recoding of the system in another framework (3jpw66).

Ce recodage du jeu de données dans un ratio proche du 1 : 1 décrit ici a le mérite de faire porter l'attention sur un problème fondamental des larges modèles. Présentés comme des copies à grande échelle, ceux-ci ne contiendraient pas une version concentrée des motifs pertinents dans l'explication d'un phénomène et ne permettraient pas, par conséquent, une théorisation subséquente des causes étant à son origine. Pis encore, certains modèles contiendraient une quantité importante d'informations non pertinentes exprimées sous la forme de variables difficiles à distinguer de celles ayant un impact réel sur les tâches auxquelles elles sont destinées. Afin d'illustrer ce second problème, une des personnes interviewées décrit le procédé d'ajustement appliqué à un modèle ainsi que sa conséquence :

> They're taking a deep model of a billion variables and they're literally slicing the data structure and throwing parts of it away, and then testing […] its predictive accuracy. In many cases, the predictive accuracy doesn't go

> down when you randomly choose parts of the training network to throw away (inyw7l).

Si les performances ne décroissent pas, cela peut signifier qu'une quantité importante de variables distribuées à travers tout le modèle sont impertinentes quant à la résolution du problème de classification auquel ce dernier est destiné.

2.4 Déploiement des systèmes d'IA et (in)interprétabilité

Une autre catégorie ou parcelle de ce problème plus large de l'interprétabilité se manifeste lors du déploiement des modèles d'IA hors des laboratoires. Une fois intégrés à des milieux professionnels, ces modèles génèrent un grand nombre d'interactions impliquant des acteurs issus de milieux ne se limitant pas au monde technoscientifique[21]. Ceux-ci expriment leurs préoccupations dans un registre différent où ce sont alors les difficultés rencontrées dans la génération *post hoc* d'interprétations portant sur l'action des modèles qui est à la source d'un malaise plutôt que le problème épistémique pour ainsi dire « pure » que cela pose.

Évoluant principalement dans le champ des sciences sociales, plusieurs personnes interviewées ont narré des récits portant sur des cas de déploiement concrets, qu'ils ne soient pleinement réalisés ou seulement à l'étape de la planification[22]. À titre d'exemple parmi ces récits se trouve ce type d'affirmation voulant qu'« étant donné qu'on utilise plutôt des algorithmes qui sont liés à l'apprentissage automatique et aux réseaux de neurones, on ne sait pas trop comment on arrive à un résultat » (j6b3dt). Les difficultés rencontrées lors de l'interprétation des actions des modèles en viennent à nuire aux efforts de traduction et de négociation nécessaires avec ceux responsables de gérer le risque au sein des milieux professionnels. Les problèmes surgissent, par exemple, lorsqu'il s'agit de « justifier au département légal » (j6b3dt) le déploiement de tels systèmes. Un scénario similaire peut se produire dans le milieu de la santé[23] où, puisque des processus vivants sont au cœur des interventions, « il y a toujours des exceptions, il y a toujours

[21] La plupart des cas rapportés par les interviewé.es lors des entretiens concernaient le domaine juridique et le secteur de la santé.

[22] Il est intéressant de noter ici que même si le déploiement ne s'est pas encore produit et que les systèmes d'IA n'existent qu'à l'état de projet, la version incarnée du malaise épistémique causée par ses déficiences en termes d'interprétabilité est rapportée par les sujets qui se projettent nécessairement dans différents scénarios impliquant des dysfonctions.

[23] Le déploiement de ces technologies dans d'autres secteurs tels que les ressources humaines, à travers la constitution de plateformes de recrutement, et l'éducation, avec l'intégration de robots conversationnels dans les cursus universitaires, provoque des problèmes similaires d'acceptabilité dans la mesure où ils bousculent des rituels sociaux établis et troublent les méthodes d'évaluations qui les accompagnent.

des problèmes, il y a toujours un niveau de marge d'erreur. Donc, justement, si ça ne fonctionne pas, on est pas capable, après ça d'expliquer ce qui s'est passé » (glv2br). Cette incapacité à expliquer les résultats produits par les systèmes d'IA « même s'ils sont vraiment impressionnants » (glv2br) empêcherait d'établir, dans certains cas, un niveau de « confiance suffisante » qui favoriserait leur déploiement. Le second trait partagé par l'ensemble de ces récits concerne le type d'action performée par les systèmes d'IA visés. Dans chacun de ces cas, il s'agit de systèmes performants une assistance à la décision. Par exemple, dans le domaine de la santé, un traitement particulier pourrait être recommandé à un patient ou bien encore, dans le domaine légal, des conseils juridiques pour être prodigués par une IA. Comme le rappelait une des personnes interviewées, un des problèmes rencontrés concernant ce type de système, « c'est qu'ils sont parfois utilisés dans des contextes de prise de décision, où les personnes qui s'en servent ne les comprennent pas suffisamment et n'ont pas les moyens de les questionner » (gk3nxq). Exposé de cette façon, l'enjeu de l'interprétabilité n'est pas seulement situé au sein des modèles, mais recoupe également le niveau de compétence de leurs utilisateurs. Les difficultés rencontrées par les usagers lorsqu'ils tentent d'interpréter les sorties de ces systèmes d'IA peuvent les conduire à construire des représentations plus ou moins erronées à propos de leur fonctionnement. Celles-ci, à leur tour, apparaissent amplifier les risques associés aux potentielles dysfonctions des systèmes d'IA.

2.5 Tentatives de résolution des controverses liées à l'interprétabilité

Les problèmes reliés à l'interprétabilité des modèles, qu'ils soient situés en amont ou en aval de leur déploiement – *ad hoc* ou *post hoc* –, sont aujourd'hui visés par un grand nombre d'initiatives. Tentant tour à tour de les éliminer ou à tout le moins de limiter les inconvénients qu'ils créent, ces initiatives scientifiques et/ou corporatives se constituent progressivement en tant qu'on sous-champ spécialisé de l'IA. Comme l'affirmait une personne rencontrée : « It's kind of an interesting ecosystem. You get a black box system and we cannot trust it. So then we create a whole parallel industry of auditing the black box system » (exm9d1). Les stratégies déployées par les acteurs de ce sous-champ spécialisé sont les suivantes : la construction de représentations permettant d'expliquer simplement les opérations internes des systèmes d'IA, la réduction de leur champ d'action et la simplification de leur fonction afin de limiter l'impact des problèmes associé à l'interprétabilité ou encore la construction d'un discours critique invalidant la nécessité de l'interprétabilité dans certains contextes de déploiement.

Près de ces questionnements, certaines des personnes rencontrées dans le cadre des entretiens œuvraient au sein de l'*IA explicable* (XAI). Ceux-ci considèrent qu'une des conditions nécessaires pour le déploiement et l'acceptation des systèmes d'IA est de faire en sorte que leur fonctionnement soit aisément compréhensible par leurs usagers. Cet objectif est ainsi décrit par l'une des personnes interviewées : « The basis of explainability is building the representations that can carry on a dialogue with people who want explanations all the way from the schoolchild to the AI expert and everything in between » (inyw7l). Pour les praticien.nes de l'XAI, ces représentations agiraient à la manière d'un support permettant la construction d'un discours intelligible pour le plus grand nombre, excédant ainsi le pré carré des expert.es. Toutefois, si la fonction de ces représentations semble consensuelle, leur nature exacte demeure l'objet d'un débat ; s'agit-il de visualisations des processus internes à l'œuvre dans le système d'IA ou de justifications des décisions prises par le système qui sont produites *ad hoc*, c'est-à-dire détachées des processus internes et générées au terme de leur action ? Pour certains individus, une solution logicielle impliquant l'XAI doit à la fois offrir une forme de soutien à la compréhension et représenter ses actions : « to be explainable, any AI system or process has to develop an ancillary representation. It must represent what it does » (inyw7l). Si de tels supports au dialogue doivent correspondre aux véritables processus internes des systèmes d'IA plutôt que d'offrir une représentation entièrement synthétique reproduisant un raisonnement humain fictif, les promoteurs du XAI sont confrontés à certaines difficultés inhérentes à la taille des modèles, mais aussi au type d'information qu'ils contiennent. S'il est facile de produire des représentations extraites des modèles entraînés sur des contenus visuels, il en va autrement pour d'autres types de données, selon l'une des personnes interviewées : « Games don't have a nice structure to them like an image does in terms of how should you represent […] the knowledge inside it in a consistent way. […] That's a big problem. […] This is similar for how like natural language has this issue too » (a9a3ir). Dans la mesure où les systèmes d'IA spécialisés dans le traitement du langage naturel sont parmi les plus répandus et utilisés par le grand public – la popularité de robots conversationnels tels que ChatGPT en constitue un exemple probant – l'élimination des barrières à l'interprétation par l'introduction de représentations des processus internes reste un projet un laborieux dont l'issue est incertaine.

Plutôt que d'affronter directement cette part d'inexplicabilité, résultat d'un déficit de compréhension et d'une absence de représentations adéquates, certains experts plaident plutôt en faveur d'une simplification des systèmes d'IA. Une des personnes

interviewées favorise cette voie où un système d'IA se compare de manière analogique à celle d'un capteur simple :

> If you're treating it as a sensor, like it's providing one input on a very small scale that is […] relatively trivial to know when it's acting up, then it might not matter as much. Again, I'd prefer it to be transparent, but realistically, it's such a small input, we're not going to be checking it every time (p9efse).

Cette neutralisation partielle du problème de l'interprétabilité grâce à la réduction du champ d'action d'un outil d'IA pourrait bien être une stratégie valable, à condition cependant d'atomiser la chaîne technologique dans laquelle il s'insère. Le caractère à la fois précis et limité d'un tel outil est ce qui en faciliterait alors l'évaluation et qui le rendrait simple et sûr au point d'évacuer le besoin de comprendre le détail de son fonctionnement. Toujours selon cette personne :

> The way that we get around this right now essentially is by making sort of a very niche, small area of AI application, such that it doesn't actually matter. […] We do not necessarily need to know why, because it's a single step in a long line. And, theoretically, it's been validated for that specific step (p9efse).

Cet argument reste convaincant tant que le regard demeure rivé sur chacun de ces déploiements pris isolément. Ceci dit, si l'ensemble de ces interventions impliquant l'IA est considéré globalement, comme faisant partie d'un *assemblage* technologique, c'est l'effet combiné de tous ces systèmes dits simples qu'il s'agit en l'occurrence d'étudier. Les synergies pouvant se déployer à travers les ramifications du réseau tissé par des objets techniques employant l'IA et l'accumulation d'erreurs mineures à travers celles-ci peuvent encore ici conduire à certaines dysfonctions systémiques. Les stratégies déployées afin d'éliminer le doute au sein de ces chaînes technologiques pourraient même conduire à des erreurs plus substantielles dans la mesure, entre autres choses, où elles mettraient un terme à la discussion et à un sain scepticisme chez les spécialistes du monde technoscientifique (Ananny et Crawford, 2016 ; Amoore, 2019).

La prise en considération de la nature insécable de l'objet technique – il doit être saisi dans la totalité de son réseau (Akrich, 2006) – conduit à considérer une autre stratégie énoncée lors des entretiens qui, davantage globale que celle-ci, vise à remettre en question la nécessité même de l'interprétabilité. Employant la conduite

automobile comme métaphore, l'une des personnes interviewées s'oppose à l'idée d'une compréhension largement distribuée à travers la société de ces systèmes :

> C'est un peu courte vue, parce que si on descend dans la rue et qu'on pose la question à tous les gens avant qu'ils montent dans leur voiture de savoir comment ça se fait, qu'elles les emmènent du point A au point B, je parierais pas mal qu'il y en a un bon paquet qui ne vont pas être capables de nous donner une explication satisfaisante (gk3nxq).

Effectivement, d'exiger de la part de non-experts une compréhension de systèmes techniques complexes tels ceux impliquant de l'IA peut sembler excessif. Toutefois, il est nécessaire de préciser cette critique et d'évaluer dans quel contexte l'interprétabilité est absolument nécessaire. Comme l'affirme un.e second.e interviewé.e, il est primordial que les spécialistes émanant du monde technoscientifique, les ingénieur.es et les informaticien.nes ayant contribué à leur conception, à tout le moins, comprennent le fonctionnement de ces systèmes :

> I would actually push back against that you have to be able to explain. I think it depends on the specific application of the machine learning model. Experts needing to understand why a decision was made is not something that I would disagree with (p9efse).

Même si cette position apparaît comme une évidence, la littérature relève plusieurs cas où ces mêmes systèmes ont connu des dysfonctions ou se sont tout simplement comportés de manière inattendue. Il s'agit ici de saisir qu'une fois déployés à l'extérieur des laboratoires, les systèmes d'IA interagissent avec un monde chaotique recelant une grande part d'imprévus, ce qui vient complexifier l'action de ces systèmes. De plus, ceux-ci ne seront pas seulement manipulés par les expert.es responsables de leur conception, mais par différentes catégories d'usagers dont le niveau de qualification dépasse parfois celle d'un non-initié et parfois non. Comme le soulève l'une des personnes répondantes :

> Ce qui est plus embarrassant c'est surtout la prise de décision à partir de systèmes dont on ne connaît pas les tenants et les aboutissants. C'est-à-dire que le problème de ces systèmes, c'est qu'ils sont intégrés à des tableaux de bord de prise de décisions, soit dans la sphère publique, pour des personnes

> qui vont avoir un impact sur les politiques publiques, soit dans les entreprises, pour des personnes qui vont avoir un impact sur le développement ou l'arrêt ou la mise en place de tel ou tel système. Quand on utilise un outil et qu'on se base dessus pour la prise de décision, il y a un minimum vital (gk3nxq).

2.6 Dialogues intra, inter et extradisciplinaires

Tout en étant une discipline caractérisée par l'hyperspécialisation de ses acteurs scientifiques, l'IA ne peut exister qu'à travers un dialogue impliquant une grande diversité d'acteurs qui permettront de la financer, lui offriront des contextes de déploiements et tenteront de déployer un cadre réglementaire afin d'orienter son développement (Latour, 1987 ; Hoffman, 2017). En l'occurrence, une forme de « contre-expertise » mobilisant des acteurs issus principalement des sciences sociales vient s'ajouter à la discussion publique tant sur le mode de la critique que de la facilitation et ce, comme pour en constituer un second pôle ayant lui-même des délimitations plus ou moins définies. Le trouble et les tensions entourant l'interprétabilité des modèles d'IA semblent ainsi avoir un effet non seulement sur la qualité des différents discours des spécialistes, mais également sur la manière dont ceux-ci sont reliés entre eux, de même qu'à l'ensemble de la société plus généralement.

2.6.1 Glissements terminologiques

Une des manifestations externes identifiables de ce trouble de la signification et de l'interprétation se situe déjà au niveau du langage même mobilisé par les spécialistes des sciences computationnelles. Plusieurs personnes rencontrées ont fait état d'une forme de glissement dans la terminologie employée pour décrire leurs activités de recherche, comme le remarque celle-ci, ces termes apparemment différents deviennent dans la pratique interchangeables :

> AI, generally speaking, is like statistics without rigor. Like, that's functionally what it is. […] Deep learning, that's sort of what got kick-started in 2012, according to my understanding. But also you can call it AI. You can call it machine learning (p9efse).

Cette reprise de l'IA par la statistique semble avoir redirigé le programme scientifique de cette discipline vers le traitement des mégadonnées. Comme mentionné auparavant,

plutôt que de poursuivre les ambitions énoncées lors du sommet de Dartmouth qui visaient originellement la reproduction de compétences cognitives humaines, l'exploitation des mégadonnées accumulées depuis l'expansion du Web a redirigé les efforts des scientifiques voulant développer justement ce paradigme dit « connexionniste ». Comme l'a souligné une des personnes interviewées, une forme d'opportunisme économique semble avoir contribué à cette redéfinition de ce qu'est l'IA :

> AI as the rebranding of predictive analytics… […] That's bullshit, right? That's just people choosing a new name for something that they're already doing because they think it's catchy and because they can get more funding for it (3jpw66).

Si ce « *rebranding* » de la recherche permet de requalifier le domaine de l'analyse de données, il convient alors d'y voir un glissement dans la signification du terme IA correspondant à une transformation de son programme scientifique. À ce glissement terminologique s'ajoute aussi une certaine cooptation de nouveaux termes provoquée par différentes institutions universitaires situées au cœur de la recherche en IA. Celles-ci, selon une des personnes interviewées, encouragent l'adoption d'une terminologie se renouvelant au fur et à mesure qu'elle apparait s'éloigner du projet historique de l'IA :

> They pointed me to a reference to a very large group, including all undergrads, graduates, and everybody working at the so-called foundations of AI lab at Stanford. […] The lab was formed two years ago. They coined the term foundation models. They don't know enough logic to reason their way out of a wet paper bag. Frankly, that doesn't mean they're not smart people, but that word, foundation model, actually appeared in the draft of the European Union AI regulation a couple of weeks ago, that's scary for me when you talk about nomenclature (inyw7l).

Au-delà de la transformation du sens rattaché à certains concepts établis, autrement dit, vient s'additionner à un rythme soutenu de nouveaux termes qui sont encore à définir. Comme l'affirme l'une des personnes rencontrées : « The challenge with artificial intelligence is that the speed at which terminology changes have rapidly accelerated » (inyw7l). Cette accélération de l'évolution terminologique de la discipline correspond en partie au rythme auxquelles sont produites les publications scientifiques responsables de leur introduction. En abordant cette problématique durant le cours d'un entretien, une

des personnes interviewées la relie à l'effervescence promotionnelle caractérisant les discours entourant l'IA :

> The acceleration of the hype about AI accelerating the invention of new terms that don't even attempt to do the scholarly work to connect them. […] So I seem to spend half my time dispelling misconceptions that arise historically from the reinvention of new names (inyw7l).

Les problèmes causés par d'éventuels glissements terminologiques n'auraient pas l'ampleur qu'ils ont acquise si le milieu de la recherche n'était pas confronté au phénomène du battage médiatique. Comme le remarque avec justesse la personne dans l'extrait précédent, c'est surtout ainsi l'effet conjugué des discours promotionnels et de certains manquements en termes de perspective historique qui provoque ces troubles dans le champ de l'IA.

2.6.2 Conversations des spécialistes au sein du champ technoscientifique

Au-delà des problèmes terminologiques affectant directement le langage employé pour conceptualiser ce qu'est l'IA, les dialogues impliquant les spécialistes du champ technoscientifique portent eux-mêmes la trace d'une tension à propos de la définition de leur discipline. Les individus interviewés ont rapporté lors des entretiens la persistance d'un clivage au sein de cette communauté qui rappelle les débats entre le clan des soignés (*neats*) et celui des débraillés (*scruffies*)[24] qui ont précédé le changement de paradigme récent introduit par les succès des techniques associées à l'apprentissage dans le cadre de compétitions telles qu'ImageNet en 2012. Cette continuité des tensions historiques au sein du champ vient en partie contredire le récit d'une domination totale de l'apprentissage profond rapporté par la littérature des STS récente (Cardon et coll., 2018). En effet, la persistance d'un courant de recherche minoritaire tirant sa source de l'IA symbolique a été remarquée par certaines personnes. L'une d'entre elles décrit une relation tendue où le groupe minoritaire est ignoré par la majorité connexionniste : « I just wouldn't characterize it as a dispute because people are doing their own thing and

[24] Les termes *neats* et *scruffies* renvoient à une fracture méthodologique dans la communauté des chercheurs en IA. (Hoffman, 2017) Identifiée dans les années 1970 au sein de la communauté spécialisée dans le traitement du langage, cette division opposait un camp priorisant la logique formelle, les *neats*, contre un second, les *scruffies*, qui employait une variété de méthodes ainsi qu'un important volume de données avec un soucis moindre concernant la rigueur mathématique. Cette division s'est en partie prolongée à travers les camps opposés de l'IA symbolique et de l'IA connexionniste.

they're not really paying any attention to the other people who are also doing something called AI » (p9efse). Cette remarque démontre la persistance d'avenues de recherche indépendantes de l'apprentissage profond. Toutefois, l'opposition entre ces courants est remise en question par d'autres participant.es qui partagent plutôt une vision où une forme d'hybridation soit en cours :

> Deep learning is really good but here's a bunch of limitations. […] Symbolics methods, like, traditional AI people, don't have these limitations. We try to combine it such that we fix the limitations or ameliorate the issues. So that's my understanding of sort of how the traditional AI folks have shifted (p9efse).

Contredisant le récit d'une fin annoncée de l'IA symbolique, les propos des interviewé. es ont indiqué la continuité de son programme de recherche à travers des initiatives minoritaires. La persistance de ces pratiques semble s'exprimer à travers une demi-intégration où certains acteurs réussissent à combiner les héritages croisés de la recherche en IA alors que d'autres font le récit d'une forme d'isolement, les tenants à l'écart des principaux circuits de financement.

La présence de voix discordantes dans le domaine de l'IA n'est que rarement rapportée à l'extérieur du cercle formé par les spécialistes. L'incompréhension d'un sous-groupe envers l'autre, même si elle est avérée, a tendance à être minime comme l'affirme l'une des personnes interviewées : « ils ne se comprennent pas trop, mais ils se comprennent autour des promesses autour des bénéfices autour des risques » (j6b3dt). C'est cette volonté de préserver les sources de financement qui, souvent, fait taire les critiques ou, du moins, contribue à les atténuer, et ce, même s'il se pourrait qu'à long terme, l'absence de critique provoque une crise dans le milieu de l'IA. Comme l'affirme l'un des individus rencontrés :

> It looks like people are afraid that this pattern is going to continue, that there's going to be a third AI winter, and then their cushy jobs and the corporate sponsorships are going to dry up. I don't think that's a narrative that people shouldn't buy into as much as they do. In some ways, it's a self-fulfilling prophecy. If you cut off the possibility of constructive criticism of your research, then that's the kind of move that would make that kind of thing happen (c3l8ej).

Cette figure de l'hiver est régulièrement rappelée dans le milieu de la recherche en IA pour désigner les périodes maigres, notamment en termes de financement. Généralement

déclenchés par une accumulation de promesses irréalistes, les hivers se produisent au terme d'un cycle où la recherche fait l'objet d'une capitalisation et d'une spéculation soutenues. Aussi, toujours selon cette personne, les débats plus profonds impliquant le cercle des spécialistes en IA mériteraient d'être mieux et plus largement exposés afin de permettre une critique plus franche et qui pourrait, de ce fait, venir assainir le milieu de la recherche.

2.6.3 Interdisciplinarité : IA et sciences sociales

Si les rapports entretenus par les spécialistes du champ technoscientifique de l'IA peuvent paraître ambigus, ceux qui se tissent entre ces dernier.ières et les représentant.es des sciences sociales le sont tout autant. Toutes les personnes interviewées ont fait mention de la nécessité d'impliquer cet autre groupe de scientifiques afin d'intégrer une expertise supplémentaire dans le développement et, surtout, le déploiement de systèmes d'IA. Cependant, une fois la déclaration de principe initiale dépassée, les entretiens révèlent que la collaboration interdisciplinaire tant souhaitée par les participants ne se produit pas ou rarement en réalité. Un des échanges menés lors des entretiens illustre cette situation :

> If you're making the claim that something is socially beneficial, then you're setting society. And that is the purview of the sociologists. Theoretically, from an academic perspective, you can't make "set claims" if you have no expertise in science and technology studies, social anthropology (p9efse).

Bien que la personne interviewée fasse de preuve d'ouverture et parle de la nécessité d'impliquer des représentants des sciences humaines lors d'un cas de déploiement, c'est la figure des divisions disciplinaires qui est ici la plus saillante. Les différents champs scientifiques sont présentés comme des silos étanches et, idéalement, une expertise correspondante doit être mobilisée afin d'intervenir dans un champ donné. Les expertises seraient sollicitées sur un mode parallèle, mais une production scientifique pleinement interdisciplinaire, c'est-à-dire sollicitant simultanément les corpus de savoirs appartenant à différentes disciplines, semble difficile à mettre en place selon plusieurs personnes interviewées.

Les difficultés causées par cette conception de la science opérant en sous-ensembles disciplinaires plus ou moins étanches se manifestent sous la forme d'une mécompréhension,

voire d'une méfiance à l'égard des corpus de savoir situés à l'extérieur du champ où un acteur évolue. Comme l'affirme une des personnes interviewées : « people in computer science departments are not used to doing cross-disciplinary or interdisciplinary research and not very used to taking seriously expertise from outside their own walls » (c3l8ej). Qu'en est-il alors d'une personne experte qui tenterait de franchir la frontière de son sous-domaine spécialisé afin de s'aventurer sur le champ liminaire de l'interdisciplinarité ? Ce geste pourrait être interprété comme étant transgressif et celui ou celle qui serait tenté par l'interdisciplinarité pourrait se faire taxer d'amateurisme, comme le révèle une des personnes participantes :

> There's this bandwagon effect of people who maybe choose not to invest energy to understand the technical scientific basis of AI, but can get on the bandwagon. And I offer to sociologists that they're amateur sociologists and amateur computer scientists too (inyw7l).

Ces paroles peuvent sembler dures, mais elles ont le mérite de rappeler qu'il est difficile de cumuler le niveau de connaissances adéquat afin de performer au sein de différentes disciplines scientifiques. C'est la rareté de ce type de spécialisation qui nuit aux efforts de recherche interdisciplinaire selon une des personnes rencontrées :

> Un autre problème, à la fois technique et social, c'est l'adéquation entre les compétences humaines et la complexité des modèles. C'est-à-dire qu'aujourd'hui, il y a très peu de gens dans le bassin de gens qui sont assez compétents pour comprendre ce qui se passe (gk3nxq).

Selon cette personne, peu d'individus se situeraient à l'intersection des champs scientifiques nécessaires pour agir efficacement, et ce, entre autres, sur un plan davantage critique. Au contraire, toujours selon cette personne, c'est plutôt ce scénario qui prévaut dans lequel « l'intersection entre l'ensemble des gens qui comprennent très bien ce qui se passe et l'ensemble des gens qui n'ont pas grand-chose à ficher des enjeux sociologiques et éthiques est beaucoup trop grosse » (gk3nxq). Cette perception d'un manque généralisé d'intérêt pour les questions sociales chez les chercheur.es en IA s'explique en partie par l'inscription des personnes rencontrées dans le cadre de cette étude au sein de courants *minoritaires*, que ce soit sur le plan du financement que des intérêts de recherche[25]. Il est probable que la

[25] Voir l'introduction du rapport quant à cette question relevant de l'échantillonnage.

position adoptée par ces spécialistes les rend plus disposés à un discours ayant des accents critiques visant leur propre discipline.

Ces divergences opposant les sciences sociales et les courants majoritaires de la recherche en IA ont été rapportées par les interviewés tant sur le plan de la formation, donc sur une temporalité située en amont de la carrière des chercheurs, que dans des contextes de déploiements. Afin d'expliquer ces divergences majeures, un des intervenants appartenant au milieu de la recherche en IA critique les carences culturelles et scientifiques causées par l'hyperspécialisation de ses collègues :

> C'est souvent des gens très pointus dans leur domaine, qui sont très forts, mais qui n'ont pas de pratique de l'inférence qui ne se repose pas sur un socle culturel de connaissance des sociétés (gk3nxq).

L'élimination de cours appartenant aux « humanités » au sein des cursus universitaires en informatique a été rapportée par plusieurs personnes interviewées. Des cours en sociologie, histoire ou philosophie des sciences parfois offerts ont été supprimés des programmes de formation afin de laisser plus de place à des contenus se rapportant directement au tronc commun disciplinaire. Pour l'une des personnes, cet appauvrissement de la formation en science computationnelle expliquerait certaines lacunes rencontrées chez les praticiens de l'IA :

> Je pense que la responsabilité des acteurs académiques est de s'assurer que ces questions-là sont enseignées, discutées parce que je pense que toute une nouvelle génération va arriver sur le marché de l'emploi pas nécessairement avec ces questions en tête. Donc c'est pas parce qu'ils vont pas s'y intéresser, c'est juste qu'on leur aura jamais parlé de ces questions-là (zfr5j8).

Comme le souligne la personne rencontrée, un manque de connaissance générale en sciences sociales nuit au développement d'une réflexion à propos de ce champ et contribue à la mésinterprétation du rôle que peut jouer la catégorie de scientifiques s'y rapportant. Cette incompréhension conduirait même les expert.es provenant du champ technoscientifique de l'IA à percevoir les intervenant.es des sciences sociales comme de potentiels pourvoyeurs d'acceptabilité sociale. Une des personnes rencontrées résume ainsi cette perception : « Vous autres, vous allez nous aider à faire en sorte que ça va être accepté socialement des gens. […] Comment on va faire en sorte que les gens vont

adhérer à la technologie » (j6b3dt) ? Ces propos rapportés où la personne interprétait le point de vue d'ingénieur.es et informaticien.nes suggèrent qu'il existe une forme d'instrumentalisation des sciences sociales lorsqu'elles sont mobilisées dans les projets de déploiement impliquant des systèmes d'IA.

Il apparaît dès lors que l'interdisciplinarité souhaitée par certain.nes est limitée, car elle n'implique pas la construction d'un savoir commun tirant sa source des deux ensembles disciplinaires. À l'inverse, une véritable approche interdisciplinaire favoriserait l'émergence d'un savoir ayant une plus grande portée critique, car nourri des regards croisés que chaque discipline porte sur l'autre. Cette récupération d'un discours favorable à l'interdisciplinarité a été soulignée à grands traits par une des personnes provenant du monde technoscientifique de l'IA rencontrée :

> Il faut que les visions se confrontent, que tu acceptes de confronter ta vision avec celle du sociologue ou du philosophe […] et que tu lui fasses comprendre ta vision et que lui te fasse comprendre la sienne, ce qui va être pertinent, sinon, comme tu dis, c'est que du baratin de dire : « oui on doit être interdisciplinaire parce que ça va garantir […] et on se retrouve avec des déclarations de Montréal qui mettent de l'avant des éthiciens et des éthiciennes et des partenaires et des discours de co-construction. What the heck! Pour moi, c'est pas opératif (gk3nxq).

En somme, cette critique des institutions montréalaises de l'IA vise un discours où une terminologie propre au domaine de la recherche interdisciplinaire semble détournée de ses objectifs premiers. Si l'interdisciplinarité tant souhaitée ne se réalise pas, c'est fort probablement par manque de connaissances croisées chez les spécialistes du champ technoscientifique et les représentants des humanités. Si ces derniers ne possèdent pas d'une base de connaissances suffisante en informatique, leurs vis-à-vis, de manière symétrique, disposent d'un savoir insuffisant à propos des fondements des sciences sociales.

3 DEUXIÈME CHAPITRE — Le régime de l'upscaling ou la matérialité et la technicité de la recherche en IA

3.1 L'upscaling et son régime de production

L'histoire des sciences computationnelles est traversée par des jeux d'échelles, autant au niveau du *matériel* que du *logiciel*. En témoigne la capacité à rapetisser la taille des composantes matérielles comme les puces et les interfaces de manière inversement proportionnelle à l'accroissement de leur puissance de calcul et aux quantités d'opérations diversifiées pouvant être réalisées dans un logiciel (Gaboury, 2021 ; Tinnel, 2023). Aussi, la recherche en intelligence artificielle n'échappe pas à cette logique : comme mentionné en introduction de ce rapport de même qu'au premier chapitre, le champ est actuellement caractérisé par la popularité des approches connexionnistes qui font la promotion des technologies de réseaux de neurones et d'apprentissage machine (*machine learning*) « souples », « flexibles » et « adaptatives » engendrant une forme particulière de *mise à l'échelle* par rapport à la taille des modèles (Cardon et coll., 2018 ; Roberge et Castelle, 2021).

Afin de contextualiser les jeux d'échelles à l'œuvre dans la recherche en IA, comment ils apparaissent dans les entrevues et ce qu'ils viennent signifier en termes de controverses quant à la technicité même du champ, ce chapitre s'inspire plus particulièrement de l'article provocant *The Bitter Lesson* de l'informaticien Richard Sutton (2019), lui-même considéré comme l'un des promoteurs canadiens des approches connexionnistes. Ce dernier affirme sans trop d'hésitation que les méthodes d'accroissement de performance ayant recours à la pure puissance computationnelle, c'est-à-dire à l'augmentation du nombre de *processeurs graphiques* (GPU) lors de l'entraînement du modèle, assureront toujours de meilleurs résultats que celles qui ont recours aux programmations humaines, comme les systèmes experts par exemple. Pour Sutton, l'accroissement de la puissance de calcul et, par le fait même, du nombre de paramètres des modèles ne peut que conduire à un gain de performance. Cette logique

d'accroissement de la taille des modèles pour développer de manière corrélative la performance de ces derniers, précède d'ailleurs les écrits de Sutton sur le sujet puisque selon le sociologue Adrian Mackenzie : « In many contemporary cases, people address the problem of generalization by seeking to increase computational power (more processors, cloud computing, clusters of computers), accrue more data or find ways of adding entirely new sources of data that augment the statistical power of the models » (Mackenzie, 2015, p. 440). L'enjeu, autrement dit, en est encore et toujours un d'*upscaling*. C'est lui qui caractérise les ambitions, possibilités et pratiques visant à accroître la taille des modèles suivant l'*idée* que cet accroissement ne peut que se résoudre en accroissement corrélatif de *performance*. Le choix d'utiliser la définition de Sutton s'explique ici par la référence au texte *The Bitter Lesson* dans la réponse d'une personne interviewée, lorsque questionnée à propos de l'*upscaling* :

> C'est très associé à The Bitter Lesson de Richard Sutton, le godfather of reinforcement learning qui est en Alberta. Il […] dit qu'historiquement, à chaque fois qu'on a essayé de faire une solution algorithmique un peu à la main la méthode mégadonnées et qu'on ne se pose pas de question, ça fonctionne mieux au fil du temps. C'est juste une question de temps. Il y a beaucoup de gens qui se protègent derrière ce claim de The Bitter Lesson. Le problème c'est que ça va aussi avec l'idéologie de la société actuelle comme quoi il n'y a pas de limite terrestre à rien. C'est-à-dire que, peut-être que si on a un budget infini d'énergie, d'information et de computation, il n'y a pas besoin de se poser trop de questions (zfr5j8).

Dans ce chapitre, il doit s'agir d'explorer de quelle manière le paradigme actuel en IA en termes de réseaux neuronaux, apprentissage machine et de ses déclinaisons participe à la mise en place d'un régime à la fois discursif et pratique particulier. Selon Michel Foucault, les « régimes » sont ce qui façonne les conduites des individus à l'intérieur des limites de ce qui est plus largement – socialement, politiquement, économiquement, etc. – compris comme étant les limites de ce qui est possible et véridique (2012). Cette « vérité » est partiale pour ainsi dire par définition, ce qui mène à la mise en tension de ces régimes et au fait que le pouvoir qui s'y exerce est le résultat d'oppositions et de conflits sur les sens de cette « vérité » (Foucault, 2001, 2012). Dans le cas de l'upscaling, la tension principale qui apparait se propager est celle entre la perspective de différents spécialistes pris individuellement, d'une part, et la promesse de résultats des grands modèles et la promotion faite par l'industrie de ces modèles, de l'autre. En effet, à lire les entrevues,

il est possible d'entrevoir que l'engouement actuel pour les grands modèles façonne des pratiques de recherche en IA en encourageant le *upscaling*, et ce, malgré certaines réticences ou même résistances de la part de certains spécialistes ou communautés de recherche ici et là sur le territoire canadien.

Rich Sutton, faisant la promotion de l'upscaling, écrivait en mars 2019 que :

> the biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. […] Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation (Sutton, 2019, p. 1).

Cette citation caractérise bien l'enjeu fondamental ici. Plus la puissance computationnelle augmente afin de soutenir des modèles toujours plus grands, meilleurs seront les résultats de ces derniers. Les pratiques d'*upscaling* seraient, selon Sutton, le moyen idéal d'atteindre de ce que des spécialistes en informatique nomment « la généralisation ». Une IA capable de généraliser serait celle capable de produire des *sorties* cohérentes dans des contextes différents, en dépit de son contexte d'entraînement et de ses paramètres fondamentaux. En effet, comme nous explique une personne rencontrée afin de démontrer l'importance de la généralisation comme trajectoire future de l'IA :

> La généralisation est vraiment un enjeu. On entraîne un système avec des données, mais est-ce que, quand il est face à une situation qui n'a jamais été présente dans les données, il [le système d'IA] va être capable de s'adapter ? Pour l'instant pas trop. Pour tout ce qui est de la programmation morale des agents moraux artificiels, la généralisation est fondamentale parce qu'on aimerait que le robot, face à une situation nouvelle, soit capable de généraliser sur des choses qu'il connaît déjà sans l'avoir déjà vu (u02uz1).

Pour certain.es, l'*upscaling* pourrait ainsi être la solution au problème du contexte : un accroissement du nombre de paramètres devant permettre au modèle dernier d'atteindre un état permettant de « généraliser ». Pour d'autres, par contre, les choses apparaissent davantage compliquées dès lors que pointe à l'horizon de cette problématique la possibilité d'une intelligence dite « générale » (AGI). En témoigne la

position d'une personne interviewée qualifiant à la fois cette généralisation du plus grand enjeu et de la controverse majeure des dix dernières années dans le milieu de la recherche :

> There was always a set of people who were interested in artificial general intelligence, but there were also a lot of researchers who were just interested in using it to solve actual problems, and I think that still exists. But a lot of the work that's going towards artificial general intelligence gets more airtime, gets more publicity than the other more mundane stuff (u02uz1).

Selon cette personne, l'*upscaling* des modèles d'IA apparait comme un ensemble de pratiques controversées parce qu'elle invisibilise les projets plus expérimentaux ou de plus petites échelles et ce, d'autant que l'engouement publicisé dans les médias est justement un engouement pour de grands modèles comme ceux alimentant l'agent conversationnel ChatGPT d'OpenAI. Ainsi, les entrevues réalisées révèlent qu'il existe effectivement des désaccords sur les pratiques engendrées par le régime de l'*upscaling*. Du moment aussi où le désaccord atteint une certaine ampleur, ce dernier caractérise une controverse au sens plus ou moins canonique donné en STS par Tomasso Venturini :

> The notion of disagreement is to be taken in the widest sense: controversies begin when actors discover that they cannot ignore each other and controversies end when actors manage to work out a solid compromise to live together. Anything between these two extremes can be called a controversy (2012, p. 261).

Si la promotion et la justification de l'*upscaling* se cristallise ici autour de Sutton, son autre et ce qu'il trouve d'écho dans une certaine critique se trouve entre autres au moment très médiatisé entourant la publication de l'article « On the dangers of stochastic parrots: Can language models be too big? » de Bender et collègues en 2021 – celui-là même faisant écho à un autre article de Bender et collègue de 2020 tel qu'invoqué dans le chapitre précédent. Cette publication est emblématique en ce qu'elle vient remettre en question les grands modèles de langage (LLMs), plus précisément la qualité de leurs *sorties* et leur impact environnementaux. Elle est aussi importante dans la mesure où elle pose la question de l'allocation des ressources dans le champ de l'IA avec notamment le congédiement d'ingénieur.es de Google, celui d'abord et avant tout de Timnit Gebru (Roberge et Lebrun, 2022 ; Marres et coll., 2024). L'idée voulant que « plus un modèle est

grand, mieux il performe », incarne ainsi la controverse ; elle met en scène non seulement des visions et des discours divergents, mais encore des moyens, des matériaux et des capacités qui sont synonymes de pouvoir.

3.2 Économie politique de l'IA

Dans cette section du chapitre, il s'agit d'explorer comment le régime de l'*upscaling* façonne la matérialité de la recherche en intelligence artificielle, à savoir l'organisation logistique des pratiques en matière d'infrastructure, de ressource et de financement nécessaires aux activités de recherche. En effet, la production de technologies à base de réseaux de neurones engendre des besoins spécifiques en termes de financement, d'acquisition de données en assez grande quantité pour assurer l'entraînement des modèles ainsi que d'accès aux infrastructures de puissance de calcul (Mackenzie, 2015 ; Pasquinelli et Joler, 2020 ; Crawford, 2021; Rella, 2023). La logique fondamentale du régime de l'*upscaling*, selon laquelle l'accroissement de la taille des modèles les rapproche de la capacité de généralisation, engendre un accroissement structurel des ressources nécessaires en recherche. À la lumière des entrevues réalisées, ce sont surtout trois éléments de cette matérialité de l'IA qui sont apparus saillants ; trois ressources qui apparaissaient comme étant justement les plus controversées : le financement, les données et la puissance computationnelle.

3.2.1 Le Canada, un écosystème scientifique centralisé

Lorsque sondés sur l'efficacité de l'écosystème de financement canadien en IA, plusieurs des interlocuteur.ices attestent un certain degré de satisfaction. Ils et elles soulignent la capacité d'acquérir un financement jugé adéquat pour la recherche expérimentale en IA. En effet, le Canada serait l'un des endroits où l'accès au financement serait le plus facile dans le monde. C'est ce qu'affirme notamment une personne interviewée : « One thing I'll note is that, in comparison to say the U.S. or the EU or U.K. Canada is by far the easiest place to get research funding » (a9a3ir). L'attractivité de l'écosystème canadien est utilisée comme argument par une autre personne, elle aussi affilié à l'un de ces instituts, afin d'expliquer son choix d'immigrer au Canada :

> Je pense que le Canada et Montréal avec le Québec, c'est effectivement un des lieux idéaux pour travailler à l'interface entre intelligence artificielle et d'autres applications. C'est pour ça que je suis venu d'ailleurs, j'avais un poste

> permanent en France et j'ai même remis en question la permanence pour un poste non permanent ici (zfr5j8).

Il y a dans ce témoignage et d'autres un engouement mélioratif pour l'écosystème canadien. Une personne chercheuse spécialisée en *traitement du langage naturel* (NLP) décrira même son expérience comme « all opportunities » (exm9d1). La réaction la plus vive aura toutefois été celle d'une personne experte en éthique de l'IA qui, en comparant le financement possible à celui disponible dans d'autres projets ailleurs, s'exclame « quand j'ai commencé à travailler là-dedans, je me suis dit "Oh là là ! là il y a un nouveau facteur à prendre en compte, c'est qu'il y a plein d'argent, quoi ! " ». Le système de financement canadien, principalement quant à la facilité à obtenir des montants, apparait ainsi à première vue adéquat pour mener des recherches en IA.

Toutefois, certaines personnes ont aussi souligné l'existence d'une dynamique de centralisation des sommes importantes de financement dans les instituts MILA, Vector et AMII par le biais de la stratégie pancanadienne en matière d'intelligence artificielle et du CIFAR – tel que vu en introduction. Cette centralisation s'effectuerait au détriment des universités extérieures à ces écosystèmes et toucherait même les fonds disponibles pour les groupes n'étant pas supportés par cette stratégie pancanadienne. Il s'agit aussi de prendre en considération la possibilité que les instituts MILA, Vector et AMII aient une facilité à conclure des partenariats avec l'industrie privée[26] dans la mesure où elle peut jouir à la fois d'une couverture médiatique favorable et d'une promotion de la part du CIFAR.

Selon une personne interviewée qui travaillant davantage sur des questions socio-centriques, le système de la stratégie pancanadienne est « politique », c'est-à-dire qu'il dépend, pour partie du moins, de jeux d'alliance entre acteurs plutôt que de la qualité des recherches menées. Elle s'explique tout en dénonçant entre autres l'exclusion de la Colombie-Britannique de la stratégie pancanadienne :

> An NSERC discovery grant is like, a really good one is what, under 100,000 a year? That's just not enough. That's the funding reality. And then, we have moved, I think, in the time I've been in Canada, more to a winners and losers model, where especially AI funding is given to a few [winning] institutions. And then the losers feel so bad. I cannot tell you the bitterness in B. C. when the big, big part

[26] Selon son rapport 2022-2023, MILA aurait conclu 119 partenariats industriels (https://mila.quebec/impact-2022-2023/).

> of funding went to the Vector Institute and to Mila. It felt very much political and people in BC felt that, you know, there was like outstanding research here and how come we didn't get any of that, right? So then you get into this weird dynamics of, you know, they're winners and losers (exm9d1).

Cette centralisation des fonds dans les trois mêmes instituts de recherche – bénéficiant de surcroît de contrats avec les grandes industries présentes dans leurs villes respectives (Google et Meta, entre autres) – s'explique en partie par la présence de chercheurs importants au sein de deux d'entre eux : Yoshua Bengio au MILA et Geoffrey Hinton au Vector Institute. Tel que vu dans l'introduction de ce rapport, le rôle de ces célèbres chercheurs canadiens en IA, récipiendaires du prix Turing et présentés comme les Parrains de l'IA (*Godfathers of AI*), est souvent utilisé comme dispositif narratif afin de promouvoir l'attractivité de l'écosystème canadien (Vincent, 2018 ; Roberge & coll. 2019 ; Dandurand et coll., 2023). Cependant, pour une personne rencontrée, ce rôle des Parrains canadiens est exagéré dans la trame narrative de l'écosystème d'innovation canadien : « We like to play up our role in providing support for basic science that NSERC has done very well, for instance, and the support of people like Geoff Hinton. » (exm9d1). Le rôle de la communauté de recherche en IA canadienne ne serait pas ainsi aussi déterminant dans la trajectoire mondiale de la technologie, et ce, contrairement aux récits promus par le CIFAR et circulant dans le milieu et dans certains médias.

Somme toute, bien que l'accès au financement par les canaux étatiques apparait convenir à la majorité des personnes rencontrées, une dynamique de centralisation des sommes versées dans les instituts MILA, Vector et AMII déplaît très certainement aux personnes qui opèrent à l'extérieur de cet écosystème et qui, de ce fait, non seulement peinent à effectuer des projets de même ampleur, mais aussi à comprendre la rationalité profonde qui se trouve derrière les allocations et arbitrages en cours.

3.2.2 Les promesses technoscientifiques comme accès aux deniers

Les entrevues ont révélé que les tensions à propos même du ou des signifiants de l'IA – ceux explorés dans la section précédente – jouent un rôle prépondérant dans la manière dont les chercheur.es font la promotion de leurs propres travaux. L'une des personnes interrogées se questionne par exemple sur la réorientation du financement en recherche « appliquée » ainsi que sur la compréhension des bailleurs de fonds quant aux projets

qu'ils financent. Comparant la conjoncture actuelle en IA avec celle de l'informatique quantique d'il y a quelques années, elle suggère que :

> NSERC's core mission is curiosity-driven research, and I would give them a sort of a B on that because they keep being pulled into funding specific programs. And I don't, I can't say I really understand what pulls them. It's not purely political, but it's not purely scientific either. And, and so I think they're wasting a lot of their money on things that they're trying to fund, on directions that they don't fully understand in the hope that they might succeed. And I mean, I'm old enough to remember when quantum computing was a big thing and NSERC poured money into that for 20 years. That resulted in basically nothing, and the success stories are being done by people like Google, who waited 20 years to think that everyone else make the mistakes and then kind of got into that game. And I sort of feel like we're doing the same thing with many of the things that go under the name of AI. It's a hot topic everyone's talking about. It's funded majorly and there's not much to show for at the end of the day (3jpw66).

Les promesses faites par les chercheur.es, les centres de recherche et les industries technologiques sont en ce sens performatives. Elles engendrent une force motrice qui est pour beaucoup une force centrifuge, à savoir que par celle-ci la recherche en IA s'attire des ressources qui sont le plus souvent à la fois symboliques et trébuchantes (Borup et coll., 2006 ; Hoffman, 2017).

Selon une personne rencontrée, la situation du financement canadien de l'IA peut se comparer à celui de l'explosion de la bulle Internet au tournant des années 2000 : « There's going to be some sorting out of these more hype like companies and research groups that are promising that they're going to solve all the human problems. Because the promises are impossible to deliver on, right? » (exm9d1). L'état de ces promesses est aussi ce qui pousse une personne à décrire le climat actuel en recherche comme prompt à favoriser la génération d'énoncés « élastiques » agissant à titre de stratégie de promotion du milieu et d'acquisition de capitaux :

> Dans une situation où c'est publish or perish, où ton financement en dépend, tu peux faire des claims élastiques un petit peu. Puis je pense qu'en IA, dans les dix dernières années, les instituts de recherche n'ont pas nécessairement

> complètement joué leur rôle de s'assurer que le discours était [...] qu'on baissait la température du discours (x04alc).

Aussi, s'ajoutant à la difficulté des chercheur.es à remplir leurs promesses, celle de trouver des débouchés canadiens pour les produits issus de la recherche. Ce problème de commercialisation n'est pas nouveau au Canada (McKenna, 2023). En effet, il est historiquement avéré que les innovations s'appuyant sur de la recherche canadienne finissent par être vendues à des entreprises américaines qui encaissent les bénéfices financiers liés à la commercialisation des produits. Cet enjeu a été problématisé par un chercheur :

> A lot of the publicity around these breakthrough points are usually, from a company, which may be American or a research group, which is in the U. S., but it is based on research done by Canadian researchers. I think that is a loss because a lot of the implementations and the thing that makes it widely known is the commercialization and all of that comes from non-Canadian researchers and companies not based in Canada (u02uz1).

Toujours selon cette personne, cette dynamique nuit à la réputation canadienne puisqu'elle occulte la renommée de sa communauté de recherche par l'exportation des innovations aux États-Unis. La recherche en IA apparaît dès lors comme nécessitant beaucoup d'injections de fonds obtenus à force de promesses soutenues sinon, superlatives, quant aux bénéfices de ces technologies, et ce, même si générant peu de retombées économiques via la commercialisation.

Cet état des lieux sur les promesses et leurs résultats, autant au niveau d'un gain de performance des réseaux neuronaux que de la commercialisation des produits, est critiqué par plusieurs personnes interviewées, dont une qui affirme : « I actually don't see very much market penetration from AI into our day to day lives. » (3jpw66) et, plus encore, que nous vivons un moment de « rebranding » de l'analytique prédictive comme de l'IA « because they think it's catchy and because they can get more funding. » (3jpw66). Deux autres chercheur.es ont exprimé dans des termes similaires l'importance de séparer la « hype » de la recherche et des produits concrets ou utiles[27]. Finalement, une autre

[27] Ici par exemple en disant « from the scientific point of view, the challenge is to separate the, the hype from the actual rigorous research. And there is a lot of hype » (exm9d1) ou là « now I would say that the primary challenge is separating or maybe it's better to say identifying where the value of these things actually are and sort of not being obfuscated by sort of hype cycles and bad actors trying to convince people to spend money that they shouldn't spend » (a9a3ir).

personne souligne les risques quant aux impacts à long terme des promesses et de l'engouement sur le financement de la recherche, rappelant l'ambiguïté de la position d'un.e chercheur.e financé.e par des fonds publics :

> Quand est-ce que tu commences à bullshitter ? La raison d'être en tant que chercheur, c'est de douter, pis de pas avoir toute la vérité de ce qui va se passer par la suite. Pis, t'as le droit en tant que chercheur de dire : « Ouais ben regarde je travaille là-dessus, ça se pourrait que ça puisse faire ça. Ça se pourrait que ça puisse faire ça ». Puis, tu vas l'écrire ton hype parce que tu le veux ton financement, parce que c'est publish or perish mais pour ça, il faut que tu aies du cash puis que tu sois capable de produire des résultats. Fait que si t'es dans une situation où t'es capable de créer un narratif qui te permet d'avoir plus de financement, une self-fulfilling prophecy, tu auras peut-être plus de chance de livrer ce qui est là. Mais en même temps, vu que t'es scientifique, tu sais pas ce qui peut arriver par la suite, tu peux y aller avec des verbes au conditionnel puis extrapoler, mais si t'étais dans le privé puis, que tu faisais le même pitch la personne elle serait là « attend minute là », je veux dire tsé : « Il est en train de… qu'est-ce qu'il veut me vendre ? Est-ce que c'est pour ses propres intérêts ? » Quand t'es dans un institut de recherche, t'es chercheur, t'as comme une protection supplémentaire ou t'as une imputabilité de moins de livrer la marchandise ou à tout le moins tsé tu as une certaine neutralité apparente. Puis je dirais que, de façon générale, c'est ça notre job de dire : « Attention, ça s'en vient. Pis, il faut regarder tous les aspects par rapport à ça. Pis, je te tiens au courant du développement de cette technologie-là, parce que ça va avoir un impact dans ta société, donc d'un point de vue législation, d'un point de vue impact économique, c'est extraordinaire ». Tout ça fait qu'il faut que tu te le crées ton hype il faut que tu t'informes en tant que société civile de ce qui s'en vient, mais ça reste quand même qu'il peut avoir de l'enflure parce que c'est aussi quand même dans ton intérêt (x04alc).

3.2.3 Les craintes comme nouvelles stratégies de financement ?

Le régime de l'upscaling a engendré au tournant de l'année 2022 une autre controverse – si davantage médiatique que celle-là comparativement au *Parrot* – autour des impacts des grands modèles d'IA et ce qu'ils peuvent venir signifier en termes de contrôle et de perspective d'avenir pour rien de moins que l'humanité. Cette dernière fait irruption à la

suite de la publication d'une lettre ouverte par le *Future of Life Institute*[28], co-signée par plusieurs acteurs gravitant autour de différents écosystèmes de l'IA, médiatiques, industriels et académiques. Ces derniers réclamaient un moratoire de six mois sur l'entraînement des grands modèles plus puissants que le GPT-4 d'OpenAI[29] (Future of Life Institute, mars 2023). Si ces questions ont fait couler passablement d'encre, rares sont encore les analystes de ce qui en a été dit et pensé dans les communautés de recherche. Lorsqu'interrogés sur cette lettre, ainsi que sur plusieurs déclarations médiatiques de signataires comme Yoshua Bengio, Geoffrey Hinton, Elon Musk et Sam Altman, ce sont de fait plusieurs interlocuteur. trices qui associent ces actions et discours à une stratégie de financement ou de centralisation des ressources.

Plutôt que de générer des promesses nourries à d'hypothétiques impacts positifs de l'IA, les personnes rencontrées estiment assister à la création, par certains acteurs, d'une forme de « panique morale » autour de ce qui est encore une fois d'hypothétiques et spéculatifs impacts sociétaux dits « catastrophiques ». Ces mêmes acteurs qui ont fait carrière sur le développement de l'IA et qui s'inquièteraient désormais des impacts de leurs technologies viennent maintenant à se positionner en fournisseurs de solutions pour régler cesdits problèmes (Chartier-Edwards et coll., sous presse). Interrogé à propos de l'explosion médiatique autour de la question des risques existentiels, un chercheur affirme que « These sorts of people, even people like, you know, Geoffrey Hinton, these people stand to earn a lot of money if they can convince people that AI is something to be scared of, and they should pay them to be less afraid of it » (a9a3ir). Pour certains, il y aurait en effet un avantage de concentration économique à se positionner à la fois comme fournisseurs d'IA et de solutions pour prévenir les risques posés par le déploiement d'une IA annoncée comme « rogue ». Il est intéressant de noter que les solutions mises de l'avant par ces acteurs, comme Hinton ou Bengio, ne s'attaquent pas à la taille des modèles et vont plutôt dans le sens du régime de l'*upscaling* en continuant d'encourager la création de grands modèles (Bengio & coll., 2023 ; Nakonechny, 2024 ; Chartier-Edwards et coll., sous presse). La mitigation des risques passerait en outre par la création de modèles plus performants dans lesquels les marges d'erreur seraient réduites.

Cette stratégie du moratoire pour prévenir l'émergence d'une IA malveillante servirait à la fois de justification pour augmenter le financement en IA, mais aussi pour intensifier

[28] Voir l'introduction du rapport ou son troisième chapitre pour les développements concernant le Future of Life Institute.
[29] Accessible à l'adresse suivante : https://futureoflife.org/open-letter/pause-giant-ai-experiments/ (consulté le 30 septembre 2024).

la centralisation des capitaux et du pouvoir dans les mains de ceux qui en bénéficient déjà. Comme l'explique une personne interviewée, « when you put a moratorium, that just means that the early adopters will benefit because they are ahead of the game and nobody else can join, right? So, no, I don't see a solution [à la concentration des capitaux et du pouvoir] other than, I don't know, break up all the big tech companies and make them smaller » (exm9d1). Dans cette situation, les *early adopters* réfèrent aux industries qui exercent déjà un certain pouvoir dans le monde de la recherche grâce aux capitaux, infrastructures de calcul et capacités de captation de données qu'elles possèdent. Selon cette personne, le moratoire serait insuffisant pour rétablir une forme d'équité dans l'accès aux ressources pour le milieu de la recherche et empêcher l'émergence d'une hypothétique IA malveillante.

En somme, le régime de l'*upscaling* apparait contribuer à la pression déjà ressentie par les chercheur.es pour acquérir du financement. Il s'agit ainsi de retenir des entrevues que, bien que le Canada soit un pays où l'accès au financement reste facile, la capacité d'accéder aux montants suffisant pour faire de la recherche sur les grands modèles semble réservée aux instituts MILA, Vector et AMII qui bénéficient du support de la stratégie pancanadienne en matière d'intelligence artificielle. Plus encore, un certain tournant appliqué du système de financement canadien encourage une forme d'économie de la promesse, dans laquelle les chercheur.es en viennent à exagérer les prouesses des systèmes d'IA et par le fait même, les attentes sur les impacts de leurs projets afin de sécuriser des montants d'argent. Finalement, l'arrivée dans l'espace public de discours à teneur eschatologique est reçue comme une manœuvre stratégique de certains acteurs leur permettant d'acquérir encore plus de financement, et ce, en se positionnant comme fournisseurs de solutions à d'hypothétiques problèmes engendrés par l'IA. Il est intéressant de mentionner à cet effet que le MILA s'est vu octroyer une somme de 21 millions de dollars par le gouvernement québécois pour encourager la recherche en IA socialement bénéfique le 13 avril 2023, 22 jours après l'appui de Bengio au moratoire sur la recherche en IA[30]. La centralisation du financement dans les instituts soutenus par la stratégie pancanadienne apparait ainsi comme relativement bien problématisée par les interlocuteurs rencontrés. Toutefois, la nécessité d'effectuer des recherches entre autres sociologiques sur les grands modèles ou encore la nécessité d'accroître exponentiellement leur taille – et les coûts affairant à ces activités – n'apparaissent que peu ou pas remise en question.

[30] Accessible à cette adresse : https://mila.quebec/fr/nouvelle/le-gouvernement-quebecois-octroie-21-m-a-mila-pour-encourager-la-recherche-en-ia (consulté le 30 septembre 2024).

3.3 Données

3.3.1 Les bonnes données pour le bon modèle

La logique expansionniste du régime de l'*upscaling* influence empiriquement la quantité de données nécessaires au développement de modèles et parallèlement, les pratiques d'acquisitions de ces données. Une personne spécialiste en gouvernance des données insiste sur l'importance de cette ressource dans le milieu de l'IA en affirmant que « the desire to get access to more and better data was driven by big data analytic and those are the same needs that are driving AI as well » (es6a2). Cette augmentation, guidée par la prémisse que plus de paramètres, équivaut à une meilleure performance, inquiète une des personnes rencontrées justement quant à la performance de ces grands modèles :

> I'm also concerned about the number of parameters that have to be set for these models to work well. And, and the number of parameters is approaching the size of the data in some cases. And I have philosophical doubts about whether you're doing anything but purely empirical fit when you build models based on that kind of data (3jpw66).

Cette affirmation est intéressante, car elle remet en cause la logique même du régime de l'*upscaling* qui voudrait que la taille accrue des modèles garantisse leur performance. Ici, la performance d'un modèle ne semble pas nécessairement garantie par sa taille, mais plutôt par la qualité des données sur lequel il est entraîné : « beaucoup des problèmes liés à l'IA sont liés au fait que les jeux de données ne sont pas bons » (kxzwsf). Il faut, autrement dit, s'intéresser à ce que les entrevues révèlent sur la quantité et aux types de données convoitées, mais aussi sur les pratiques d'acquisitions.

Une multitude de questions, de commentaires et de critiques a été formulée à propos du rôle des données et des pratiques d'acquisition de ces dernières dans la construction des modèles. En effet, la quasi-totalité des personnes interlocuteur.ices a affirmé que la majorité des enjeux et problématiques engendrés par l'IA s'expliquent par des problèmes situés au niveau des données et de leur collecte :

> I did spend a lot of years in industry and a lot of the problems often amounts to the fact that the data is not in the form that is that we want it to be in. Sometimes it's proven that the source is not clear and that's not helpful because you need

> to know what context the data was collected in before you use it for certain models (u02uz1).

Les pratiques de collecte qui séparent radicalement les données de leur contexte seraient nuisibles lors de la création de modèles, car elles rendent difficile l'utilisation de ces mêmes données pour entraîner des modèles devant exécuter des tâches spécifiques. Plus encore, « you have to start with data, you know, you have to figure out what data you're using and how clean it is and what it isn't » (exm9d1), nous explique une personne chercheuse, lorsque questionnée sur les enjeux techniques les plus importants dans la construction des modèles. Elle souligne que les résultats produits par ces derniers dépendent directement de la qualité des entrées de donnée. Afin de garantir la performance des modèles, une autre personne affirme l'importance d'un jeu de données cohérent avec les objectifs à réaliser : « ensuring that you're actually training on a training data set relevant to the problem, right? So you're not wasting time later having to clean, right? » (a9a3ir). L'importance de la relation entre jeu de donnée et objectif du modèle explique l'insistance des interlocuteur.ices quant à la nécessité d'avoir *les bonnes données pour le bon modèle*. La question de savoir ce que constitue les « bonnes données » se décline alors en deux temps, soit un premier sur les *méthodes d'acquisitions* puis, un second, sur le *contenu* de ces dernières.

Si l'extraction de données est une étape fondamentale et nécessaire pour mener à bien la construction des modèles (Mackenzie, 2015 ; Pasquinelli & Joler, 2020 ; Crawford, 2021), les entrevues indiquent que les techniques d'acquisition apparaissent comme un objet de controverse dans les communautés de recherche en IA. Par exemple, une personne explique que les quantités toujours grandissantes de données massives pour entraîner de grands modèles ont pour conséquence le recours à des méthodes d'acquisitions qui ne respectent de la vie privée des internautes :

> So, one of the issues was finding enough data to be able to train deep learning models, and then there are sort of consequences of the ways that was done for privacy. Also, as a result of needing very large data sets, there are starting to be environmental concerns emerging as well (c3i8ej).

La nécessité de toujours avoir plus de données et d'obtenir ces dernières plus rapidement afin de rester compétitif dans l'écosystème d'innovation, engendre un extractivisme qui tente toujours de repousser les frontières de la disponibilité publique pour acquérir des données de meilleure qualité ou traitant de sujets bien particuliers.

L'absence d'information sur les contextes de collecte pose un enjeu quant à l'éthique de la recherche, et ce, entre autres puisque les données sont souvent récoltées sans le consentement des individus (Zuboff, 2020 ; Crawford, 2021). Plus encore, les personnes rencontrées soulignent qu'à l'absence de consentement, s'ajoute l'absence de régulation au Canada sur les manières de se procurer les données, notamment quand elles concernent des communautés marginalisées. Cette situation inquiète :

> Comment tu fais en tant que société pour t'assurer que tu réponds aux besoins de toutes les communautés, pour que tu sois capable de les capturer [les données]de manière éthique ? C'est pas l'IA qui va trouver sa solution à ça. C'est vraiment la société qui doit mettre en place une façon de collecter des données tout en étant respectueux de la vie privée. Je pense que c'est plus vers les politiques publiques que c'est que tu seras en mesure de le faire (x04alc).

La personne citée à l'instant insiste : les pratiques d'acquisition de données doivent être encadrées par l'État, car les solutions techniques sont inefficaces. Ce qui est inquiétant, c'est que dans certains cas, la performance des algorithmes peut dépendre de la capacité à éroder la vie privée lors de la collecte pour acquérir un avantage compétitif sur les autres joueurs dans la composition des jeux de données. C'est ce qu'affirme cette chercheuse spécialisée en éthique et vie privée en ligne : « so privacy is different from security, right? Like, all of these companies purposefully conflate the two, like secure data privacy and data security. But these algorithms don't work unless you violate the privacy of everyone » (p9efse).

Le manque de formation dans les cursus sur les méthodes d'acquisition des données qui respectent la vie privée pourrait, selon une autre personne expliquer l'attitude délinquante ou maladroite des industries dans la collecte des données : « But privacy becomes an issue when it comes to data access in terms of security, first of all, and then it's an also an issue because the practitioners don't necessarily or are not necessarily trained with, privacy in mind when they're using the data or processing the data » (u02uz1). Par ailleurs, une forme d'asymétrie est soulignée quant à la capacité des acteurs industriels de collecter des données massives en ligne versus celle des acteurs pratiquant la recherche dans un contexte académique :

> Il y a [des] controverses sur le plan sociétal bien sûr, souligne une personne interviewée, quant à savoir comment gérer la gouvernance des données et des

> modèles, surtout en présence de gros acteurs qui peuvent se permettre de dépenser des quantités astronomiques en apprentissage et de l'autre côté, une recherche académique qui n'a pas forcément les mêmes moyens (zfr5j8).

Cette asymétrie favorise les grands joueurs industriels en encourageant une dynamique de concentration des ressources. En l'occurrence, les chercheur.es du milieu académique qui n'ont pas les moyens de récolter massivement des données en ligne et d'organiser ces dernières d'une manière adéquate peuvent acquérir des ensembles « prêt-à-utiliser » organisés par des compagnies de courtages de données. Ces dernières organisent leurs ensembles à partir de données achetées chez les grandes compagnies, notamment Google. Cette dynamique est décrite par une chercheuse experte en éthique et gouvernance des données :

> there are companies that have data warehouses and data lakes, and you have to be able to access them to get the kind of data that you want in the form that you want. There's often a data delivery issue and issues on whether you're able to get it with the speed that you need and whether you have the memory to run it (u02uz1).

Cette dynamique a entre autres comme conséquence que la recherche académique en IA vient à dépendre de plus en plus de l'industrie technologique pour avoir accès à des ensembles de données assez larges et organisés pour travailler. Ainsi, l'extractivisme numérique, dont les conséquences sociales multiples ont déjà été largement abordées sous un angle critique par les sciences sociales (Zuboff, 2020 ; Crawford, 2021; McKelvey, 2021), se retrouve également à être problématisé par les practicien.nes de l'IA dans le corpus des entrevues. D'une part, les personnes observent le repoussement des limites de la vie privée en ligne afin d'acquérir toujours plus de nouveaux types de données ; de l'autre, ils constatent la dépendance progressive du milieu académique face au milieu industriel afin de faire de la recherche sur de grands modèles à moindre coût.

3.3.2 Les limites du déchétarisme numérique

Les entrevues révèlent tout autant l'existence de problématiques sur le contenu des données. En effet, ce dernier a un impact direct sur les performances et résultats des modèles. C'est ce que souligne d'emblée une personne spécialisée en *machine learning* :

> Là, on n'est plus dans la forme, mais dans le contenu des informations qui sont utilisées, et si on voulait avoir une approche ultra-clean, ultra-respectueuse, ça veut dire qu'il faudrait se contenter d'utiliser des contenus qui sont du domaine public et qui, d'autre part, ne contiennent pas d'informations personnelles. Il faudrait que les informations ne puissent pas être recoupées avec des bases de données extérieures pour apprendre des choses privées sur les personnes (gk3nxq).

Une partie au moins du problème est à dire que l'introduction de données de piètre qualité produit des résultats n'ayant pas plus de valeur qu'une pile de déchets ; d'où ce qui enclenche la recherche pour des données propres, bien étiquetées. Tel que l'explique l'une des personnes rencontrées :

> Data is sort of what lies at the foundation of these systems. The challenges early on have really been getting those data sets, curating them, annotating them and having the computing power to actually deal with. Now that, you know, computing power isn't as much of an issue, data sets and data quality and processing are, I think, are still very important and will sort of continue to be into the future (qnkqsf).

Une forme de hiérarchisation-monétisation des données apparait s'installer ; celles qui sont facilement accessibles sont dévaluées au profit de données difficilement accessibles, par exemple des données biomédicales issues du domaine de la santé. Comme expliqué précédemment dans les entrevues, acquérir des données précises en dehors des bases déjà constituées soit par des organismes publics, des acteurs industriels ou des courtiers de données est difficile pour le milieu de la recherche. Cette difficulté d'accès a un impact sur la capacité de la communauté de recherche à produire des modèles pouvant traiter d'enjeux précis qui nécessitent justement des données bien structurées et organisées. C'est ce qu'explique une personne qui fait de la recherche sur des modèles d'IA en santé publique :

> The data is definitely a problem because sometimes, especially when people work in health care. There's like one data set they have to get, and it's always going to be MIMIC[31] and therefore MIMIC kind of just tells you what kinds of

[31] L'acronyme MIMIC désigne une famille de jeux de données destiné à la recherche en santé (Johnson et coll., 2023) Voir : https://www.nature.com/articles/s41597-022-01899-x

> problems you can solve. They only have like, three outputs of classes you can try to predict, and they have sometime series data, but they don't have, like, longitudinal data for a patient, like, over a long period of time. And, so yeah, the first step is like what data is available and often, that really restricts what you're able to do (u54r89).

La problématique apparaît comme récurrente : la valeur du contenu des données en fonction de la difficulté posée par les capacités d'acquisition place les acteurs industriels de grande envergure dans une position de dominance par rapport aux autres acteurs, puisqu'ils sont déjà en possession de données organisées. Cette position leur permet de tirer avantage des écosystèmes de recherche ouverts et de la collaboration académique, selon un répondant :

> Ce n'est pas nécessairement l'algorithme ou les modèles qui sont développés qui ont de la valeur. C'est les bases de données. Tu peux avoir différentes entreprises, différents chercheurs, ils vont vouloir collaborer ensemble à essayer de développer les meilleurs modèles possibles parce que y a aucune création de valeur nécessairement au modèle qui est là. Fait que t'as l'ouverture, t'as les publications. Mais en fait les gens qui peuvent faire de l'argent sur des modèles d'IA c'est ceux qui sont assis sur des données de qualité. Comment tu régules ça ? En ce moment, ce n'est pas pour rien, que ce soit Amazon ou Google peut se dire, « moi, là, je te finance accoté pis, regarde, publie c'est pas un problème là je vais continuer à financer, whatever, t'as toute la liberté du monde ». Parce que la compétition ne se fait pas pour développer le meilleur algorithme. Oui peut-être un peu, mais de façon générale, c'est open source parce que dans le fond, l'important c'est les données (x04alc).

Ce qu'il serait possible de nommer les « miettes numériques » ne suffisent plus pour combler les besoins d'entraînement dès lors que l'objectif de la recherche n'est plus de gagner des concours de performance en reconnaissance d'image, par exemple, mais plutôt de développer des modèles aux applications concrètes entraînés sur des bases de données spécialisées correspondant à des secteurs d'activité précis. À l'image du commentaire précédemment mentionné sur le tournant en recherche appliquée des financements gouvernementaux, l'intensification des maillages entre l'académie et l'industrie participe aussi à ce tournant appliqué. Les recherches académiques doivent servir à développer

des modèles d'IA en fonction d'usages concrets et précis. Ce tournant appliqué favorise encore une fois les acteurs industriels qui peuvent injecter des sommes importantes d'argent en recherche afin de sécuriser des données particulières :

> L'appât du gain, c'est les données. C'est comment faire pour accaparer le maximum de données ? Financer la recherche en IA, c'est tout à l'intérêt de ces entreprises-là qui sont assises sur ces coffres de trésor-là de données. Sur les montants de financement industriels, y'a deux choses là-dedans : je trouve que les montants d'investissement qu'il y a en recherche surtout dans un domaine de recherche qui est logiciel, ce qui fait que ça coûte fuck all parce que c'est juste des cerveaux puis après un paquet d'ordinateurs, je pense que c'est du jamais vu avec un si peu nombre d'acteurs aussi qui sont qui sont globaux. Oui ça crée quand même une situation qui est exceptionnelle, qu'on n'a pas vu avant là (X04alc).

Le besoin de données de qualité et en grande quantité, surtout si les méthodes d'acquisition visent à respecter la vie privée, engendre de nouveaux coûts pour la recherche en IA et agit ainsi à titre de justification pour un accroissement des financements ou, du moins, un rapprochement des modes opératoires de l'académie et de l'industrie. Aussi, à terme, il apparait que ce régime de l'*upscaling* favorise tendanciellement la recherche à teneur industrielle plutôt que purement académique alors que la trajectoire actuelle tente de développer des applications pratiques et monétisables pour et par les technologies d'IA génératives. Ces modèles appliqués, souvent sous la forme de LLMs, nécessitent des données massives qui doivent cependant être bien organisées et nettoyées afin d'éviter la génération *de sorties* biaisées, racistes, discriminatoires, etc. De plus, les besoins en données plus précises pour alimenter ces modèles semblent encourager des pratiques d'extraction qui enfreignent la vie privée des internautes, ce qui va à l'encontre de principes éthiques en recherche. Les chercheur.es en contexte universitaire qui veulent développer de grands modèles se retrouvent pour ainsi dire à la remorque des industries puisqu'ils et elles n'ont pas accès aux mêmes moyens en matière de collecte et d'organisation des données. Ils et elles doivent acheter aux industries ou courtiers en données des ensembles « prêts à utiliser » lorsque les ensembles fournis par les organismes publics ne suffisent pas à leurs recherches. Ce qui est de ce fait révélé par les entrevues, c'est une dynamique sous-jacente de dépendance progressive du milieu académique de la recherche en IA envers les industries, principalement quant à l'accès aux ressources nécessaires à la recherche.

3.4 Puissance computationnelle

3.4.1 L'accès aux infrastructures de calcul

Les deux sections précédentes ont exploré les impacts du régime de l'*upscaling* sur les montants de financement et les quantités de données nécessaires à la construction de grands modèles. Ici, il s'agira de se concentrer sur la dernière ressource dite « technique » mentionnée dans les entrevues, c'est-à-dire la puissance computationnelle. En effet, le gain de performance des réseaux de neurones est largement attribuable à l'utilisation de *processeurs graphiques* (GPU) et à la disponibilité de mégadonnées, conditions nécessaires à l'entraînement de modèles comportant toujours plus de paramètres (Rella, 2023). Ce tournant est d'ailleurs l'incarnation exacte de la logique de Sutton dans *The Bitter Lesson* mentionnée ci-haut. Les GPUs, comme composante matérielle de l'IA, sont désormais indispensables à quiconque veut effectuer de la recherche ou de la construction de modèles : « If you want to do any kind of deep learning, you need GPUs. And even now, a lot of medium-sized universities still struggle to get enough GPUs to support training in courses but also in research » (u54r89). La littérature documente déjà les conséquences de cette nécessité : une course aux acquisitions qui favorise les industries et acteurs ayant accès aux grands circuits de financement alors que les chercheur.es en contexte universitaire deviennent dépendants de ces mêmes joueurs afin de mener à bien leurs recherches sur de grands modèles (Togelius et Yannakakis, 2023). Les entrevues révèlent l'existence de cette même dynamique en contexte canadien, à savoir qu'encore une fois, les acteurs industriels sont ici favorisés et que les universités peinent à obtenir les infrastructures de calcul nécessaires. Une personne rencontrée illustre cette dynamique en nous donnant en exemple une situation problématique entraînée par la difficulté d'accès aux infrastructures :

> When I was at U of T [University of Toronto], I was teaching a course on natural language processing. I couldn't get enough GPUs for my students to do anything. With GPUs, I wanted them to train some models for speech recognition and like, The U of T is a well-resourced university by Canadian standards. And I couldn't get enough GPUs for someone to maybe run one iteration of fine-tuning and that's it. And that's not counting even if they had a bug or if they have to do it again, they couldn't anyway. So, the GPUs are a problem (u54r89).

En effet, l'incapacité d'obtenir un nombre suffisant de GPUs pour faire fonctionner un LLM dans un contexte d'enseignement implique nécessairement que certaines activités de formation sont difficilement réalisables – ce cours précis avait d'ailleurs dû être annulé. Le déficit de moyens dans les universités ne peut qu'accentuer la tendance de dépendance des universités envers les industries. À cet effet, il est intéressant de noter que les acteurs industriels visés par les répondants sont généralement issus des États-Unis, notamment OpenAI, Google, Microsoft et Amazon.

Cette dynamique de concentration de la puissance computationnelle est d'autant plus dénoncée qu'elle confère aux industries le pouvoir d'établir plus ou moins directement les bancs d'essai (*benchmarks*), c'est-à-dire les normes d'évaluation de performance des modèles. Ces mêmes seuils inspirent et guident éventuellement la recherche universitaire, ce qui est inquiétant selon cette experte : « I think that a bunch of people are also worrying about how to, to build a state-of-the-art model. You need the kind of compute power that very few places in the world now have. And so how much of what counts as state-of-the-art results is driven by industry » (c3l8ej). Cette dynamique participe au régime de l'*upscaling* en confondant la taille des grands modèles avec la performance de ces derniers par l'établissement justement des bancs d'essai. Il s'agit ici d'une situation analogue à celle de l'acquisition de données énoncée plus haut ; le contrôle industriel sur les GPUs et les bancs d'essai incite fortement les membres des communautés de recherche universitaire à accepter les modalités de la recherche partenariale et de la science ouverte afin de se rapprocher des industries pour avoir accès aux infrastructures de calcul dont ils ont besoin. En effet, une chercheuse prend d'ailleurs le temps de souligner le nombre de partenariats industriels dans le champ de la recherche en IA, ainsi que l'influence de cette proximité grandissante entre l'industrie et l'université sur la recherche en elle-même :

> A lot of people have noted that, just about everyone working in AI has some kind of corporate partnership or is getting money from the industry. And while people don't like to believe that they're influenced by that sort of thing, or that their integrity would be marred in some way by getting money for something, we know from psychological research that people are influenced by this. *So whether we like to believe that or not, there's good reason to think that industry is changing the direction of research and that it's pretty clearly the reason why they're sponsoring research.* If we want to have the kind of products that would actually fulfill this promise of this sort of brighter future through AI, we need research money without strings attached (c3l8ej).

En sommes, ce milieu de la recherche en IA se caractériserait par un accroissement des partenariats publics-privés afin de permettre aux chercheur.es en contexte universitaire de travailler sur de grands modèles. Cette asymétrie dans l'accès est dénoncée par une autre chercheuse qui s'inquiète des effets à long terme sur les trajectoires futures de la recherche en IA tant que les industries continuent d'agir comme forces structurantes sur le milieu :

> I think it's a big problem that these are impressive models that OpenAI has published. It was only possible by them because they had billions of dollars of investment to get them the infrastructure needed to build those. There's no way that could have happened at a university. I think that is a very big issue and a big hindrance to the proper development of AI because now it's really only the big companies that are able to have these big developments like this. Once these kinds of large models have been developed and published, who's going to pay attention to somebody's little logistic, logistic regression model that they have in their corner, right? (u02uz1)

Une autre personne abonde dans le même sens en liant les enjeux de financements et d'accès à la puissance computationnelle comme facteur explicatif de la force structurante de l'industrie sur la recherche universitaire en IA : « you're more likely to get access to Google grants to buy yourself 100 GPUs, if you agree with their vision and mission and that affects the research that can be done and affects society as well » (p9efse). Au sujet de la puissance computationnelle, les personnes interrogées rapportent, dénoncent et critiquent même avec une certaine virulence cette dynamique de concentration de ressources au sein d'industries le plus souvent américaines. La présence de projets industriels et la publicité faite aux grands modèles semblent minimiser l'importance de la recherche fondamentale en IA[32]. La porte de sortie pour ceux et celles n'ayant pas les moyens d'acquérir des GPUs et ne voulant pas ou ne pouvant pas se prévaloir de partenariats industriels serait de recourir à des modèles préentraînés (parmi lesquels figurent les modèles de fondation) (Bommasani, R. et coll., 2021) afin d'économiser en nécessité de puissance de calcul. Plutôt que de devoir entraîner un modèle à partir de rien, le recours à un modèle de fondation permet au chercheur ou à la chercheuse de se concentrer sur des efforts d'ajustements qui nécessitent moins de puissance computationnelle. Cette solution peut cependant prendre la forme d'un « cadeau empoisonné », puisque les *modèles de fondation* comme GPT ou BERT sont déjà des productions industrielles :

[32] Deux personnes interviewées ont donné en exemple le travail expérimental pouvant être accompli grâce à des *toy problems*.

> The availability of pre-trained models has helped in that. I can use transformer models without having 10 GPUs at my disposal, which I do not have. That's one technical and financial help. With pre-trained models I only have to do fine-tuning. on them, but that's also letting somebody else make the decisions, right? Companies so far have been making pre-trained models available. We don't know how long that'sgonna last. OpenAI is already doing that. They're saying that "we're only going to sell our products to Microsoft first and everybody else gets our second-best model" (exm9d1).

Le recours aux *modèles de fondation* ne règle pas le problème de la dépense aux financements, données et infrastructures industrielles pour faire de la recherche en IA. La concentration des GPUs dans les industries et plateformes notamment américaines agit comme un facteur intensifiant de la compénétration entre académie et industrie et la solution de recourir aux *modèles de fondation* n'est qu'un prolongement de cette problématique. Ceci dit, une autre solution serait de ne pas faire de la recherche sur de grands modèles, ce qui pourrait paraître comme contreproductif dans l'obtention de financements vu l'engouement actuel pour ce genre de recherche.

3.4.2 Un problème écologique

Encore un autre enjeu souligné par les répondant.es dans les entrevues porte sur la délinquance écologique engendrée par le nombre de GPUs requis pour faire fonctionner les modèles d'IA. En effet, la dépense énergétique croissante peut être attribuée aux grands modèles et pratiques d'*upscaling* (Crawford, 2021; Bender et coll., 2021). Cette incompatibilité des pratiques actuelles d'*upscaling* avec plusieurs engagements écologiques quant à la sobriété énergétique est soulignée par cette personne qui affirme que « le problème c'est que c'est en total désaccord avec justement les défis écologiques vers lesquels on s'avance, quoi ? C'est-à-dire qu'avec l'*upscaling*, on va vers une décadence écologique totale » (zfr5j8). Une autre personne abonde dans le même sens, affirmant qu'il y a un écart entre la consommation énergétique des modèles versus les résultats de ces derniers :

> We have to face up to the fact that it's using an awful lot of energy, right? I mean, not as much as crypto, but still there's GPUs pulling power like you wouldn't wouldn't believe. And that's all got to come from somewhere. And frankly, the, the amount of electricity that has to go into one tiny result. It's, it's

> quite alarming when you consider that you'd like to replicate this and scale it and all sorts of other things. Right. And I just don't see how that can work in the sense of power consumption (3jpw66).

Exaspéré, un énième chercheur affirme dans le même sens :

> Je ne peux pas concevoir qu'on consomme autant d'énergie pour ces modèles-là pour les résultats que ça donne versus justement ce qu'on sait que le cerveau humain est capable de faire avec si peu de puissance. Là, je trouve qu'on court à notre perte. Honnêtement, pour moi, l'enjeu technologique principal, c'est la consommation d'énergie de ces modèles-là qui sont un petit peu efficaces. Ils donnent de bons résultats, mais on ne peut pas continuer comme ça (x04alc).

Ainsi, les besoins en énergie des grands modèles ne seraient pas soutenables d'un point de vue de frugalité énergétique. Le décalage entre la consommation énergétique et les résultats produits apparaissent ainsi comme un sujet de controverse doté d'une certaine constance. La nécessité dans le champ de souvent devoir recalculer des paramètres semble aussi déranger cette interlocutrice : « Hey, ça va faire, là, les dépenses environnementales inutiles pour recalculer des machins pour la 150 000e fois. Donc, comment on fait pour avoir une approche [...] peut-être pas frugale, mais au moins raisonnable de l'utilisation de la puissance de calcul » (gk3nxq). Un manque de sérieux quant aux engagements écologiques des universités, mais aussi des industries est dénoncé de manière récurrente. En effet, une autre chercheuse s'inquiète de l'éventuelle catastrophe écologique engendrée par une propagation des grands modèles, qualifiant cette dernière de cauchemardesque :

> It's not worth the environmental costs. And the research arms race, well it used to be just research. It used to be researchers at universities. We were all trying to get a few more GPUs. But now it's companies fighting with each other for more and more servers and data centers. And, that's just it, you know, capitalist nightmare (exm9d1).

Toujours selon elle, les recherches sur les modèles de langage auraient dû rester au stade de projets de recherche plutôt que de devenir des produits industriels, notamment afin de restreindre leur taille et certains de leurs usages. Plus encore, l'accroissement des modèles ainsi que leur propagation rapide pousse une interlocutrice à souligner l'enjeu de la soutenabilité à long terme de la recherche en IA :

> Scaling up to the next generation of chat GPT implies to be multiplying the number of parameters and the size of the data set by a hundred again, to get to your next generation. Then you're in territory where it's actually not possible to do with the resources on the earth. […] And I think there's fairly good agreement that it's not really possible to make them that much bigger than they are now. That the resources involved are just like unreasonable on a like earthly scale, so there's not much more that you can go with that kind of method (c3i8ej).

En tout et pour tout, l'engouement actuel pour les grands modèles motive une course à l'acquisition des GPUs en tant que composantes informatiques des plus coûteuses. Ces dernières viennent généralement à être monopolisées par les grandes industries américaines au détriment des chercheur.es qui pratiquent en contexte académique, peu importe leur positionnement dans l'écosystème canadien. Ceux et celles qui ne peuvent pas accéder à un nombre de GPUs suffisant afin de travailler sur de grands modèles doivent soit conclure un partenariat avec un partenaire industriel afin d'accéder à ses infrastructures de calcul, soit recourir à un *modèle de fondation* lui aussi issu de l'industrie afin de réduire la puissance computationnelle nécessaire, ou bien se tourner vers des projets de recherche à plus petite échelle. La valorisation des projets sur les grands modèles industriels appliqués occulte d'ailleurs les recherches plus fondamentales ou expérimentales en informatique en sorte que se profile ici encore une dynamique d'intensification de dépendance du milieu académique envers le milieu industriel. Au-delà de l'enjeu de l'accès aux infrastructures de puissances computationnelles, des personnes interviewées ont souligné certains problèmes d'errance écologique qui caractérise l'engouement pour les grands modèles, notamment en termes de consommation d'énergie. Autant au niveau économique qu'au niveau écologique, la configuration actuelle de la recherche en IA sous le régime de l'*upscaling* ne serait pas ainsi soutenable à long terme.

Ce chapitre du rapport, à travers les résultats d'entrevues portant sur la matérialité de l'IA, porte ainsi bel et bien sur la controversialité de l'*upscaling* autant au niveau des pratiques de recherches, des exigences en termes de ressources ainsi que des performances des modèles. Plusieurs personnes rencontrées ont bien problématisé l'*upscaling* dans les entrevues. Par exemple, l'un d'entre le qualifie de « changement d'ère » qui nuit à la recherche en contexte universitaire :

> On entre dans une ère où justement il y a les modèles fondationnels ou les large models qui demandent des compétences d'ingénierie informatique au-delà de

> la compétence mathématique et informatique théorique. Et dans ce cas-là, il y a ce problème technique qui se combine avec un problème socio-économique qui est que, le peu de personnes qui sont capables de faire ce genre de passage à l'échelle sur le marché de l'emploi sont agressivement démarchés par à peu près tout le monde et que dans ce contexte-là, le monde académique est pas forcément super attractif pour ce high qualified personnal (zfr5j8).

Les enjeux soulignés au long du chapitre seraient tributaires selon cette personne d'une fuite des cerveaux du secteur public vers le secteur privé. Ce même chercheur prolonge sa critique en affirmant que la recherche académique vit actuellement une transformation qui évacue la théorie[33] :

> La critique que je ferai c'est que le scaling est très à la mode parce que ça remplace la théorie, c'est-à-dire qu'au lieu de se poser plus de questions sur comment faire ça mieux on a juste à allonger plus de chèques et à avoir plus de GPU ou plus de data. C'est plus facile d'un point de vue épistémique de faire ça que de se poser la question de comment refondre et repenser les concepts et les algorithmes sur lesquels on travaille (zfr5j8).

La controversialité autour du régime de l'*upscaling* vient de ce fait contester la logique fondamentale de cette dynamique telle qu'énoncée par Sutton dans *The Bitter Lesson* : l'accroissement de la taille des modèles serait corollaire à un gain de performance. Cette logique même est problématisée à tous les niveaux de la matérialité de la recherche en IA. Plus de financement, de données massives et de GPUs ne garantissent pas exactement les gains de performance. Comme le note cette chercheure : « we're going to sort of plateau on how many improvements can be made to the kinds of models that, that people are sort of successful with right now » (c3l8ej). Cette stagnation probable des grands modèles reste à voir, mais une chose est certaine, la logique fondamentale de l'upscaling est remise en question par les practicien.nes interrogés.

En voulant émuler ou compétitionner le genre de recherche qui se fait dans l'industrie, les universitaires perdraient de vue la possibilité de faire de la recherche fondamentale et expérimentale. Cela illustre aussi bien un autre élément important de la controversialité inhérente au régime de l'*upscaling*. Il s'agit de la concentration des capacités à accroître la taille des modèles dans les industries notamment américaines, dont la présence en

[33] Voir le premier chapitre pour plus d'information à propos du concept de « fin de la théorie ».

sol canadien se fait de plus en plus sentir dans les partenariats publics-privés impliquant le monde universitaire. La difficulté éprouvée au niveau de la commercialisation des technologies numériques au Canada – bien que les recherches qui contribuent à la commercialisation de technologies états-uniennes soient effectuées en sol canadien par des chercheur.es canadien.nes – a été de plus mentionnée par certaines personnes interviewées. En plus des enjeux d'accès et de soutenabilité à long terme traité tout au long du rapport, il apparait important de souligner un élément fondamental de la controverse autour du régime de l'*upscaling* : en termes de *souveraineté numérique*, l'avenir de la communauté de recherche en IA canadienne tend à dépendre de plus en plus des infrastructures industrielles américaines pour mener à bien ses projets et survivre financièrement. Cet élément apparait fondamental, car révélateur d'une ambiguïté performative importante : bien que tous et toutes pointent du doigt la « *big tech* » Américaine comme, en quelque sorte, « l'ennemi » qui impose sa volonté d'accroissement de la taille des modèles comme force structurante de la recherche, la plupart viennent s'adonner à des activités partenariales avec celle-ci. Au-delà des enjeux matériels explorés tout au long de ce chapitre, l'élément central de la controverse apparait ainsi se situer au niveau de la compénétration progressive du milieu de la recherche *universitaire canadienne* avec celui des *industries et des plateformes principalement américaines* et ce que cela peut signifier en termes d'autonomie des chercheur.es canadien.nes et de leurs recherches.

4. TROISIÈME CHAPITRE — Quand les modèles d'IA quittent les laboratoires pour se répandre dans la société : entre hype, hallucinations et préjugés

Comme mentionné ici et là dans le rapport, les modèles d'IA ne sont plus aujourd'hui confinés aux laboratoires de recherche universitaires. Ils sont déployés de plus en plus rapidement dans la société et façonnent désormais simultanément les mondes sociaux, économiques et politiques. Non seulement cette situation constitue-t-elle en un risque de dommages imprévus, mais elle appelle également à une réflexion sur la manière dont l'IA devrait être réglementée. Lorsqu'interrogés sur la construction sociale et les impacts de l'IA, les spécialistes rencontrés ont apporté des réponses – de manière surprenante – plutôt homogènes en pointant vers ce qui pourrait être considéré comme une compréhension « sensationnaliste », sinon même parfois mal informée de l'IA par les publics comme l'un des principaux problèmes de l'époque. En effet, la plupart des personnes rencontrées se sont ralliées à l'idée que le fait d'associer des voix plus diverses à la conception, à la gouvernance et à la couverture médiatique de l'IA pourrait être la stratégie corrective et d'atténuation des risques la plus efficace. Dans ce troisième et dernier chapitre du rapport, les réponses des différentes personnes interviewées ont été regroupées autour de trois thèmes. Le premier explicite les différentes manières dont les discours promotionnels – entre autres ceux de ces promoteurs scientifiques et économiques – et tout l'engouement pour ce qui touche de près ou de loin à l'IA façonnent ce qui est compris de la technologie et du rôle qu'elle est appelée à jouer en société. La seconde thématique, quant à elle, souligne comment les capacités techniques de l'IA générative, une fois diffusée en masse, risquent d'accroître de manière exponentielle les problèmes de désinformation et de mésinformation. Enfin, une troisième thématique renvoie aux tensions existantes en ce qui a trait à la gouvernance de l'IA, ce qu'elle peut avoir de démocratique ou d'une science citoyenne dite « participative », ou plutôt si elle demeure la chasse gardée d'une poignée d'expert.es. À terme, il sera ainsi question d'enjeux de traduction des différentes connaissances à l'œuvre dans le façonnement

de l'IA et donc de sa construction sociale, notamment à la lumière de cette dynamique généralisée d'engouement pour l'IA ; l'objectif général étant d'exposer certains des principaux domaines de controverse et de consensus au sein de la communauté des chercheurs en IA au Canada en relation avec le déploiement et la réception plus larges de ces systèmes.

4.1 Le battage médiatique à propos de l'IA dans un monde de faible littératie technologique

Lorsqu'ils examinent les incidences sociales de l'IA, les personnes rencontrées partent souvent d'un constat commun : le public est encore peu informé de la réalité, de la nature et des capacités réelles des systèmes d'IA. Compte tenu des nombreuses craintes et promesses y étant liées, un tel niveau dit « d'alphabétisation » leur apparait comme nécessairement préoccupant. Tout d'abord, beaucoup considèrent que le public manque d'expertise en matière d'*interaction* avec les systèmes d'IA (Collins & Evans, 2002), c'est-à-dire de l'expérience et des connaissances constituées en multipliant les interactions avec ces systèmes mêmes. Ce manquement ouvre la voie au « battage » et aux promesses entourant les capacités de l'IA, ce qui le plus souvent contribue à façonner des attentes irréalistes. Pour plusieurs personnes interviewées, l'imaginaire populaire de l'IA s'apparente à une « baguette magique qui allait tout sauver » (lwillq). Selon une personne, cela est dû à « un problème d'*expectation management* où les gens prétendent que ces choses-là sont magiques et répondent à tous les problèmes » (kxzwsf), ce qui s'expliquerait au moins en partie par la couverture médiatique de l'IA. « Ce qu'on lit dans les journaux, dans les médias, c'est assez rare que ce soit en ligne [avec les grandes controverses qu'on traverse dans le domaine de l'IA] parce que c'est extrêmement sensationnaliste », est-il expliqué (kxzwsf). Tel qu'observé dans un rapport précédent de l'équipe de recherche *Shaping AI* (Dandurand et coll., 2022), l'IA est principalement représentée dans les médias traditionnels canadiens comme une puissante solution commerciale destinée à révolutionner la vie quotidienne et les flux de travail d'une manière qui éviterait toutes frictions ou presque. Cependant, cette couverture ne permet pas d'informer correctement le public des détails techniques importants – *et des limites* – de ces systèmes.

La couverture médiatique, comme l'indique une personne rencontrée, peut paraitre assez superficielle au point où « in almost every media discussion of AI, you could replace it with the word magic » (3jpw66). Ce « fossé » entre « the way that [AI is] portrayed to the

general public versus the actual details » de ces systèmes est un problème largement reconnu par les personnes rencontrées (p9efse). « I don't think the media actually understands very much of what's going on » (3jpw66), affirme une autre personne. « There's a lot of very gullible people who see something like [ChatGPT] and say 'it's wonderful, we've made a huge advance in human understanding, the world is going to change tomorrow.' And, uh, I worry about those people, many of whom seem to be journalists » (3jpw66).

L'engouement actuel pose ainsi un ensemble de problèmes. Par exemple, lorsque des controverses surgissent dans le discours public, les critiques semblent souvent mal alignées sur les réalités techniques et, de ce fait, facilement rejetées. Une personne experte déplore que « des fois les scientifiques, puis les chercheurs, puis les industries condamnent vite le grand public d'être des ignorants qui ne sont pas capables d'évaluer, disons, une technologie [ou] qui bloquent des développements parce qu'ils ont une conception arriérée des choses » (glv2br). Le problème n'est pas tant le manque d'information du public en soi, mais plutôt la pression incessante pour commercialiser les systèmes d'IA sans prendre le temps d'expliquer correctement leur fonctionnement : « c'est parce qu'ils ont été mal informés ; on n'a pas pris le temps de bien comprendre, on a lancé des choses avec urgence » (glv2br). Il existe plusieurs raisons potentielles pour lesquelles les journalistes et les reporters technologiques – malgré parfois une compréhension personnelle avancée de l'IA – parviennent difficilement à informer correctement le public sur la technicité de l'IA et ses implications sociales. Il peut s'agir, par exemple, des pressions institutionnelles et économiques des salles de rédaction sous-financées (Fenton, 2011 ; Compton et Dyer-Witherford, 2014), du sensationnalisme croissant des formats journalistiques (Marinov, 2020), ou peut-être parce que les discussions approfondies sur la technicité de l'IA ne sont pas en soi considérées comme « dignes d'intérêt ». Quoi qu'il en soit, la couverture médiatique sensationnelle ou non substantielle nuit, selon certain.es expert.es, à des discussions sérieuses qui seraient d'intérêt public. Par exemple, une personne rencontrée revient sur son « conflit avec les journalistes » pour dénoncer l'absence d'applications et de débats sur l'IA axés sur le bien public :

> I think a plausible definition of AI is anything computational that journalists can't understand. And that's the trouble, right? All the stuff that's been in the public discussion is ignorant and uninformed. And I don't know how you get voices into that space of people who know what they're talking about and can make the case for, you know, some positive sides of AI, um, given the pressure from businesses in the other direction (3jpw66).

En d'autres termes, sans une couverture bien informée alimentant une compréhension publique significative, il devient difficile d'imaginer ce que pourraient être les différentes réceptions sociales, mais aussi, les différentes formes de contestations démocratiques des systèmes d'IA – en particulier en réponse aux agendas industriels

4.1.1 Le battage médiatique et la littératie constituent-ils un risque pour le domaine de l'IA lui-même ?

Les discours sensationnels et sous-informés sur l'IA n'ont pas seulement un impact sur le grand public, ils marquent également de leur empreinte les communications commerciales et scientifiques, faisant ainsi peser des risques sur la recherche en IA elle-même. Les préoccupations sont multiples. Par exemple, la prolifération de terminologies de plus en plus vagues autour du terme générique « IA » empêche les différents acteurs de communiquer correctement les significations multivalentes de l'IA. Si certains dénoncent la « manipulation des médias », d'autres considèrent qu'il s'agit d'une tactique des industriels. S'agissant de la différence entre l'utilisation du terme « IA » par Apple et Google lors de leurs conférences de développeurs, une personne experte note que si « they both utilize the exact same technology », ils commercialisent l'IA en utilisant des terminologies différentes : « It's all about branding… the words you use depends on who you're trying to sell it to » (p9efse). Les outils et systèmes d'IA étant désormais omniprésents dans le discours public, les spécialistes ne sont pas à l'abri d'interprétations extérieures, parfois hyperboliques, des technologies qu'ils créent, y compris de la part des entreprises pour lesquelles ils travaillent. Ainsi, la fluidité terminologique ou la « tournure » a également un impact sur la manière dont les spécialistes en IA parlent entre eux. « People have had to change the way they're describing their work because the public has started to understand those words in different ways », explique une personne rencontrée (p9efse). À mesure que les différentes méthodes statistiques d'analyse des données et d'apprentissage automatique prolifèrent et sont regroupées au hasard (ou omises) sous la terminologie générale d'« IA », les communications non seulement entre « the more technical people » (p9efse), comme l'explique une personne, mais aussi entre les différents domaines d'expertise peuvent devenir confuses.

Par exemple, les experts en IA au Canada ont du mal à s'y retrouver dans les diverses critiques qui ne reflètent pas toujours les réalités techniques sous-jacentes. Comme l'explique une personne interviewée, « the acceleration of the hype about AI » a également accéléré « the invention of new terms that don't even attempt to do the scholarly work to

connect them » (inyw7l). Le problème s'illustre en utilisant l'exemple de l'IA « responsable » ou « éthique », déplorant ce qui peut être considéré comme un « bandwagon effect of people who maybe choose not to invest energy to understand the technical scientific basis of AI » (inyw7l). La prolifération des termes, des concepts et des critiques aggrave les difficultés à naviguer dans l'utilisation déjà confuse et interchangeable du terme « IA » pour parler de plusieurs méthodes de calcul distinctes. Ces incohérences, même si elles sont bien intentionnées, contribuent à créer un répertoire interprétatif qui sème la confusion parmi les publics et les experts[34].

La crédibilité et la rigueur scientifique du domaine sont aussi à risque ; elles peuvent en effet être compromises par le battage médiatique et les tactiques de recherche d'attention. Un expert évoque les effets négatifs concernant « the acceleration of ideas and nomenclature from academic science lab to industry to the public », expliquant que « the trainees and apprentices with a passion for science » tentent de plus en plus de faire avancer le domaine « in a simplistic way because they get incredibly tangled up by [the] flavor of the day » (inyw7l). À mesure que les terminologies et les concepts explosent dans le domaine public, les universitaires commencent à « se bousculer pour attirer l'attention », explique-t-il. L'un des moyens les plus faciles afin d'obtenir de l'attention est « to pick a controversial topic, invent new phrases and say, 'Oh, I've got a novel solution to that problem' » (inyw7l). Le risque que des tendances concurrentielles « au goût du jour » noient la recherche scientifique rigoureuse est également très préoccupant sur le plan méthodologique. Comme le fait remarquer un spécialiste, les étudiants en IA et en informatique apprennent surtout à coder et moins à « faire de la science » ou à rester « sceptiques ». En conséquence, explique-t-il, « people are very happy about a 0.1% improvement over a state of the art » et vont souvent vanter leurs résultats sans tenir compte d'une série de questions méthodologiques (u54r89), par exemple si l'augmentation n'est qu'un « accident » ou une comparaison injuste des modèles[35]. Ainsi, comme le résume une personne rencontrée, le défi « d'un point de vue scientifique » consiste à « to separate the hype from the actual rigorous research. And there is a lot of hype » (exm9d1).

Ce problème existe depuis des décennies, bien sûr, l'histoire de l'IA étant « characterized by a lot of over-promises and disappointments going back to that sort of

[34] Voir le premier chapitre pour des explications détaillées concernant les problèmes de terminologie et de définitions affectant le champ technoscientifique de l'IA.

[35] Voir le deuxième chapitre pour discussion approfondie portant sur les bancs d'essai et la notion de SOTA.

post-World War II period » (qnkqsf), comme l'explique l'une des personnes rencontrées. Aujourd'hui, cependant, ce problème est mis en évidence par le langage dominant du technosolutionnisme :

> The fact that AI is such an amorphous term that's not very well understood generally and inconsistently applied means, I think, that it can be used or presented in this way where it can be sort of held up as a solution to just about anything. And because we are sort of riding high on this wave of AI hype that has a tendency to sort of proliferate these solutionist promises in ways that are wildly unrealistic, right? (qnkqsf)

C'est précisément là que les conséquences des dangers du battage médiatique et du manque d'information apparaissent au grand jour. Comme l'explique une personne, « the primary challenge » consiste à identifier la valeur réelle offerte par les systèmes d'IA et à « not being obfuscated by sort of hype cycles and [when] a bad actor is trying to convince people to spend money that they shouldn't spend » (a9a3ir). L'empressement à commercialiser les produits d'IA gonfle souvent les affirmations exagérées sur les applications ou le potentiel de l'IA et alimente ce battage médiatique pour stimuler les ventes. En conséquence, sans une compréhension technique ou même contextuelle suffisante des systèmes d'IA, et avec des imaginations exagérées quant aux capacités « magiques » de l'IA, la réception sociale de l'IA, la capacité des publics à contester les applications problématiques ou exagérées et les connaissances scientifiques partagées entre les experts sont toutes affectées de manière négative.

4.1.2 Naviguer dans la vague médiatique des Parrains de l'IA

Selon les personnes interrogées au cours du projet de recherche, la source la plus importante de l'engouement pour l'IA ces derniers mois se trouve dans les débats publics portant sur les opinions divergentes des Parrains de l'IA – notamment Yoshua Bengio, Geoffrey Hinton et Yann LeCun – concernant les allégations selon lesquelles l'IA constitue une « menace existentielle ». Dans une lettre ouverte datant de mars 2023, Bengio, Hinton et plusieurs autres ont appelé à une pause de six mois dans le développement de grands modèles d'IA en raison du risque de menaces imprévues, voire existentielles, pour l'humanité (Future of Life Institute, 2023)[36]. Des appels similaires ont ensuite été lancés par d'éminents acteurs industriels tels que Sam Altman d'OpenAI et Bill Gates de Microsoft (Centre pour

[36] Disponible à l'adresse suivante : https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

la sécurité de l'IA, 2023)[37]. La couverture de ce débat a été particulièrement importante au Canada, contribuant à la prolifération du battage médiatique et des discours sur l'IA « éthique, responsable, sûre » et « socialement bénéfique » évoquée plus haut, ainsi qu'à l'état général de confusion dans ce domaine.

Pour les gens rencontrés, la question d'un éventuel moratoire s'avère controversée. La plupart des expert.es sont sceptiques quant au discours sur les risques existentiels. Les personnes rencontrées privilégient généralement une approche plus nuancée et parlent de problèmes particuliers et précis dont les risques sont « gérables » et qui ne sont pas exactement de type existentiel. Si les personnes sont au moins quelque peu préoccupées par l'automatisation des emplois et les applications militaires de l'IA, elles soulignent la longue histoire de l'automatisation du travail et les nombreuses années d'existence d'armes et d'outils militaires améliorés par l'IA pour dissiper les craintes. Les spécialistes restent généralement critiques à l'égard de l'hyperbole et de l'attention médiatique démesurée accordée aux partisans des « menaces existentielles » – attention qui pourrait, selon certains, se traduire par une monopolisation accrue de l'espace de l'IA.

L'interprétation de l'attention portée par les médias aux Parrains de l'IA est largement influencée par la dynamique entre le battage médiatique sur l'IA et la littératie sur le sujet. Partant de l'écart entre les capacités réelles de l'IA et la compréhension qu'en a le public, certaines personnes estiment que des acteurs tels que Hinton et Bengio cherchent à combler cet espace en tant que traducteurs ou intermédiaires du savoir. Comme l'affirme un expert de manière critique :

> Toute cette discussion-là, c'est encore de dire comme l'IA c'est surpuissant, on pourra pas la contrôler, ça va tous nous tuer. » C'est pas vrai, mais ça contribue à alimenter les vignettes que c'est une technologie complètement, complètement hors de contrôle, pour laquelle on doit avoir des gens spécialisés qui se faire d'intermédiaire entre nous et cette force supérieure (kxzwsf).

Il poursuit cependant en suggérant qu'un tel autopositionnement est plutôt normal dans les situations de développement de haute technologie :

> Mais fondamentalement, c'est un peu normal que les gens s'attachent aussi à ce genre de personnalité là, parce que… [pour] toutes les technologies, une

[37] Disponible à l'adresse suivante : https://www.safe.ai/statement-on-ai-risk.

> fois que t'arrives à un certain niveau, [ils] sont complètement déconnectés de la compréhension des gens normaux. Et puis c'est [présenté] comme une entité externe, omnipotent, omnipuissant et tu nécessites un prêtre qui fait la médiation entre toi et cette entité externe qui a un pouvoir extrême sur ta vie (kxzwsf).

Pour certaines personnes interviewées, ces personnalités apprécient simplement d'attirer l'attention et la notoriété, la possibilité d'exprimer leur opinion ou, plus généreusement, tentent de remplacer ou « corriger » ce qu'ils perçoivent comme un champ journalistique mal informé. Pour d'autres, « le cœur est à la bonne place » : « Je pense qu'en effet Geoffrey Hinton et Yoshua Bengio à ma connaissance, ce sont les seuls chercheurs au monde à avoir levé des drapeaux rouges ou orange pour conscientiser les gens » (lwillq). Ce n'est pourtant pas la question la plus sociologique qui soit, car, quel que soit le point de vue éthique sous lequel on les considère, les interventions médiatiques de ces traducteurs de connaissances ou de ces agitateurs de drapeaux ont un impact réel sur la manière dont la recherche et le développement en matière d'IA sont structurés.

Il est important de noter que certaines personnes rencontrées supposent que les acteurs prééminents se comportent de *manière stratégique* dans la poursuite de leur agenda – que ce soit pour le financement, l'influence ou la concrétisation de positions déjà dominantes dans le champ. Selon une personne experte, cela est particulièrement probable dans le contexte québécois dans lequel opère Bengio : « Ouais ça c'est extrêmement québécois […] avoir cette culture-là de rock star basée à Montréal, ça attire beaucoup de couvertures extrêmement positives. Mais la science spectacle, c'est comme ça que t'attires plus de subventions. C'est comme ça que t'as plus de publications » (kxzwsf). Cette stratégie d'accaparement des ressources « spectaculaires » peut également fonctionner à l'inverse, c'est-à-dire lorsque les experts se positionnent de manière excessivement défensive face aux critiques par crainte d'un nouvel « hiver de l'IA ». Un interlocuteur suppose que les chercheurs craignent que « their cushy jobs and the corporate sponsorships are going to dry up. ». Comme il l'explique, « someone like Yann LeCun, I think, is a great example of this, where any kind of even reasonable and mild criticism gets a really big reaction as though it's, you know, you're saying something terrible about AI or you're a technophobe if you criticize anything » (c3l8ej). Toujours selon cette personne, de telles réactions découlent probablement de « fights for funding going on between different camps for reasons like having notoriety or having power, you know, getting research money » (c3l8ej).

Au-delà de la politique de financement, un risque encore plus grand pour le domaine provient de la possibilité que des craintes exagérées monopolisent davantage le champ de la recherche en l'IA. La surenchère des discours sur les risques existentiels peut se traduire par une influence accrue des acteurs déjà dominants dans les processus réglementaires. Cette situation a par exemple été divulguée et discutée publiquement par le cofondateur de Google Brain, Andrew Ng, qui a affirmé en octobre 2023 que des entreprises de la Big Tech ont gonflé certains discours hyperboliques et certaines « bad ideas that AI could make us go extinct » afin de déclencher des formes plus strictes de réglementation qui, de ce fait, pourraient évincer les concurrents plus petits (Davidson, 2023). Il s'agit là d'un point de vue très répandu. Comme le suppose une personne rencontrée :

> There's people like Geoff Hinton, who should know better, who say, "Oh, this is going to revolutionize the world." But I always wonder when people talk like that, if they–how much of a hidden agenda they have? I think some of the [people] trying to clamp down on models like ChatGPT is because that would make them a monopoly of the five or six big players in the space already, and they do not want small businesses coming up with innovative extensions to those models (3jpw66).

Ces inquiétudes sont monnaie courante dans les entretiens menés. Comme le fait remarquer une autre personne, « a lot of the public debates have been with people whose primarily stake in the debate is how much money they're going to make, right? » (a9a3ir). En contrôlant le discours et la formulation des menaces liées à l'IA, les acteurs dominants exercent une plus grande influence sur l'élaboration des réponses publiques et réglementaires. « C'est encore les mêmes qui sont à l'avant-plan, puis qui sont allés chercher l'intérêt du politique, puis qui sont visibles dans les médias, donc qui sont crédibles auprès des citoyens », remarque une personne critiquant le manque de diversité des voix dans l'espace de l'IA (j6b3dt). Avec la même cohorte d'acteurs qui gagnent en couverture et en notoriété, le battage médiatique autour des capacités et des risques de l'IA devient ainsi un outil politique puissant.

Dans l'ensemble, le débat sur les « menaces existentielles » est un sujet de controverse majeur. De nombreux experts ne croient pas que les IA soient réellement aussi intelligentes et capables de causer autant de dégâts que ne le suggère le battage médiatique. Cette surenchère risque aussi d'invisibiliser de nombreuses questions sociales plus concrètes

et plus immédiates posées par le déploiement de l'IA. Cherchant à maîtriser la rhétorique, une personne revient sur la faillibilité humaine pour orienter le débat sur les risques de l'IA dans une direction plus terre à terre : « Well, I think there's a couple of areas where you'd be right to be worried, but not so much because of the AI itself. It's because of the fact that we don't know what we're doing with it, and we don't therefore know when things could go horribly wrong » (3jpw66). En d'autres termes, l'orgueil peut être plus préoccupant que les inventions technologiques. C'est pourquoi plusieurs préfèrent mettre l'accent sur les lacunes de l'IA. Un commentaire en particulier résume une grande partie de ce qui a été entendu :

> I don't think there's reason to fear that AI is going to be too intelligent, and even if it did, that it would then decide to kill us all. Like there's, there's just a lot of gaps in that argument. Um, the dangers seem to come from AI not being as smart as we think it is and using it for things where it's, you know, not up to the job. That, that's like a legitimate fear (c3l8ej).

Ce type de craintes est bien plus important aux yeux de la communauté canadienne des expert.es en IA. Comme il s'agira de le voir dans quelques instants, les problèmes de littératie en matière d'IA, de battage médiatique et d'agitation autour des menaces existentielles semblent occulter un risque social beaucoup plus immédiat lié aux déploiements de systèmes d'IA.

4.2 Automatiser la dés/més/information

Le discours sur les « menaces existentielles » a dominé les discussions publiques au cours de l'année 2024, noyant sans doute les critiques importantes sur les préjudices déjà existants de l'IA (Gebru & coll., 2023 ; McKelvey, 2023). L'un de ces préjudices, simplement évoqué dans la lettre ouverte, mais qui a gagné en importance dans les discussions publiques est l'utilisation de l'IA générative afin de produire et diffuser de la désinformation et de la mésinformation. Dans le domaine des études critiques sur la désinformation, celle-ci est généralement définie comme « all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit » (Freelon et Wells, 2020). Elle se distingue donc de la « désinformation » qui elle implique des informations erronées ou trompeuses. Les travaux existants sur l'IA et la désinformation ne s'intéressent généralement pas à cette distinction, utilisant les termes de manière interchangeable pour parler uniquement d'actions

intentionnelles. Cette absence de distinction a pour conséquence que les critiques relatives à d'importantes questions techniques et épistémologiques sous-jacentes à l'IA générative, qui pourraient potentiellement entraîner la diffusion d'informations erronées « accidentelles », néanmoins socialement préjudiciables, sont souvent négligées ou rejetées. Les points de vue des spécialistes rencontrés dans le cadre de cette étude ont mis en lumière les risques de la désinformation et de la mésinformation et la manière dont les systèmes d'IA générative peuvent les façonner.

Les systèmes d'IA générative pourraient être utilisés pour créer des contenus mensongers de plus en plus sophistiqués ou personnalisés – une question qui a fait l'objet d'une attention croissante de la part des universitaires et des journalistes préoccupés en particulier par la manipulation électorale (Hsu & Thompson, 2023; Morrish, 2023). La question a même été mentionnée dans le récent décret du président américain Joe Biden sur la sécurité de l'IA (Kertysova, 2018 ; Brkan, 2019 ; Kaplan, 2020)[38] et est citée comme l'une des principales préoccupations des pays du G7 en ce qui concerne la croissance de l'IA générative (OCDE, 2023). Dans ces discours – et de fait, dans une grande partie de la littérature existante sur le sujet –, la question est presque exclusivement abordée en évoquant un cadrage par « mauvais acteurs », à savoir la crainte que l'IA générative donne davantage de pouvoir à ceux qui ont des intentions mauvaises ou manipulatrices – y compris des États tels que la Chine et la Russie, typiquement des « ennemis » de l'Occident. Or, l'impact même de ces « mauvais acteurs » est ce qui a récemment fait l'objet de controverses. Dans un commentaire publié en octobre 2023, un groupe de chercheurs affirme par exemple que les craintes concernant l'impact de l'IA générative sur la désinformation sont « exagérées » (Simon, Altay et Mercier, 2023). En se concentrant trop étroitement sur le cadre des « mauvais acteurs », de nombreux chercheur.es négligent des facteurs importants de la technologie elle-même. En ce qui concerne le contexte canadien, les personnes rencontrées dans le cadre de la présente étude indiquent que la production de fausses informations involontaires constitue effectivement un site majeur de risques liés à l'IA en raison précisément de la nature technique de ces versions en constante évolution de l'IA générative.

Il y a ici ce qui semble être un consensus clair – bien qu'implicite – sur un risque plausible posé par l'IA générative : le fait que des publics non informés fassent usage pour ainsi

[38] Outre les références académiques, se référer au document intitulé « Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence » (Maison-Blanche, 30 octobre 2023). Disponible à l'adresse suivante : https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

dire quotidiennement des systèmes d'IA générative, et ce, malgré leur technicité toujours occulte. Si la plupart des spécialistes s'inquiètent du risque prétendument « exagéré » de désinformation par de mauvais acteurs, la question la plus pertinente et la moins explorée est ainsi celle de la production « involontaire » ou « accidentelle » de *fausses informations* par les utilisateurs ordinaires. Les spécialistes déplorent souvent la manière « non critique » dont les utilisateurs acceptent les résultats de l'IA générative comme des affirmations de vérité en grande partie non problématiques. Pour beaucoup, il s'agit là du défi sociopolitique le plus pressant pour le déploiement de masse de l'IA générative.

4.2.1 Désinformation à dessein : L'IA entre les mains de « mauvais acteurs »

À première vue, l'IA générative est un outil puissant pour les « mauvais acteurs » qui cherchent à obtenir des gains économiques ou politiques par le biais d'une publicité manipulatrice, de « *fake news* » et de la diffusion d'informations fausses ou personnalisées. Si les algorithmes de recommandation peuvent contribuer au problème, les systèmes d'IA générative sont désignés comme un nouveau risque majeur susceptible de diminuer la qualité des discours circulant dans l'ensemble de la société. La capacité de créer des contenus faux ou manipulateurs existe depuis longtemps, mais le problème posé aujourd'hui par l'IA générative, explique une personne rencontrée, est l'ampleur et la sophistication nouvelles qu'elle permet :

> Ce qui change aujourd'hui, c'est l'échelle. C'est l'échelle mondiale à laquelle on est capable de diffuser ces contenus et de les propager. Et puis l'échelle de réalisation, on peut en générer à volonté en une fraction de seconde. Et de la qualité de la fausseté. C'est-à-dire que c'est vraiment bluffant la qualité avec laquelle on est capable de générer des mensonges, que ce soient des mensonges vidéo, audio, écrits (gk3nxq).

La désinformation et la publicité manipulatrice peuvent directement influencer les actions des citoyens par le biais de contenus séduisants ou de « nudges ». L'émergence d'avatars dotés d'une intelligence artificielle et utilisés dans la publicité commerciale et politique constitue alors un alors problème relativement nouveau. Comme l'explique un spécialiste de la désinformation et de l'IA, « we can create avatars for ads, and these avatars, based on AI, they can be so sophisticated that they will use the facial expression or the emotion

that will trigger people to maybe buy a product, or maybe vote for that candidate » (2vtuf2). Dans le contexte actuel des mégadonnées et de la prolifération des profils numériques des consommateurs ou des citoyens, l'ampleur des informations disponibles sur les utilisateurs de technologies ne peut être minimisée. Comme l'indique le spécialiste « if [bad actors] want to manipulate the whole community, it's not that hard. So if you have someone that wants to influence elections, the behavior of the consumers, with AI, they have a lot of power to do that » (2vtuf2). La question va donc clairement au-delà de la crainte des « fake news » si souvent évoquée dans les discours publics et concerne plutôt le pouvoir d'influence et de manipulation que représentent des millions, voire des milliards de contenus ciblés et d'incitations affectives destinés aux citoyens et aux consommateurs sur une base quotidienne.

Comme mentionné, cette question commence à faire l'objet d'une attention croissante dans les discours scientifiques, populaires et politiques. Loin d'être un problème « exagéré », ces risques semblent très réels pour les personnes rencontrées. Pour reprendre les termes d'une personne interlocutrice, « things like nudging and, at least what some people believe happened with the Cambridge Analytica scandal, of affecting democratic processes, that kind of thing could start happening if there are even more convincing tools for convincing people of political messages » (c3l8ej). Ou, comme l'a déclaré une autre avec plus de gravité :

> I'm concerned about just the ripping apart of social fabric. I mean, I do think that generative AI is just going to contribute to misinformation and disinformation, to the degradation of knowledge, to the degradation of, you know, these things that are so important to keep our societies coherent and structured and organized (es6a27).

La question de savoir si les efforts déployés par les « mauvais acteurs » pour produire de la désinformation à l'aide de l'IA générative sont une crainte exagérée reste ainsi ouverte. Néanmoins, des exemples réels provenant du monde entier prolifèrent rapidement (Ryan-Mosley, 2023) et les expert.es canadien.nes y trouvent déjà des raisons de s'inquiéter. La plupart n'ont toutefois pas grand-chose à dire sur ce problème des « mauvais acteurs » dans la mesure où les problèmes les plus préoccupants semblent provenir des utilisateurs « innocents » de l'IA générative eux-mêmes.

4.2.2 Désinformation accidentelle : Dépasser le cadre des mauvais acteurs

Les systèmes d'IA générative présentent un risque pour les utilisateurs insouciants en raison de l'acceptation accidentelle ou non critique de résultats erronés, que ce soit sous la forme « d'hallucinations », d'erreurs d'inférence, de préjugés reflétés ou de chambres d'écho personnalisées. Ces préoccupations ont déjà été exprimées par des critiques (Gold & Fischer, 2023), bien que la question n'ait guère été prise en compte dans les discours politiques au-delà de vagues aspirations à une IA « éthique ou responsable ». Elle semble également être largement négligée dans la recherche universitaire sur l'IA et la désinformation. Les risques impliquent ici des (re)productions et des percolations d'informations fausses ou problématiques distribuées à grande échelle dans toute la société, provoquées par l'intégration rapide des agents conversationnels.

Pour plusieurs spécialistes, le problème se situe au niveau technique où les fondements des LLM les prédisposent à la production fréquente de résultats erronés et peu fiables. Les personnes rencontrées font souvent écho à de critiques similaires à l'article largement débattu sur les « perroquets stochastiques » dans lequel les auteures affirment que les LLM sont des systèmes « for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning » (Bender et coll., 2021 ; Roberge et Lebrun, 2021)[39]. Une personne rencontrée explique très clairement le fond du problème :

> Hallucinate is an interesting word because it's not actually a hallucination, right? What it is, is it's trying to maximize this reward function of trying to produce something that looks factual. And that's how it produces these hallucinations. It's trying to produce something that a human will think, a human who is a non-expert will think, looks factual, right? That's just how the reward function was set up for the [reinforcement learning] portion of that. But what they're really trying to do is they're trying to not disagree with you (a9a3ir).

Sans référence logique ou contextuelle au monde, les LLM et autres systèmes génératifs génèrent du contenu avec la meilleure précision statistique possible, mais ne savent pas nécessairement quand admettre qu'ils ne savent pas. La personne poursuit :

[39] Voir le premier chapitre de ce rapport pour analyse détaillée de cette problématique.

> You can go to a ChatGPT or any of these other large language models now and have them tell you something, and then say, “No, that’s not true.” And they’ll generally say, “Yes, you’re right. That’s not true. Here’s why.” Right? And you could do this infinitely. You could keep on telling them they’re wrong and they’ll keep on bending over backwards metaphorically to agree with you (a9a3ir).

En effet, d’un point de vue technique, les hallucinations sont normales et « inévitables » (kxzwsf) ; elles sont tout simplement caractéristiques de la façon dont les systèmes d’IA générative ont été conçus et « a fundamental part of how these systems operate » (qnkqsf). Toujours, selon cette personne, les résultats hallucinatoires ou erronés semblent « inherent [in] that these are probabilistic, autocomplete or plagiarism machines that just inherently make stuff up. That’s what they do. They have no understanding of the world. All that they can do is predict the next word or token in some sequence » (qnkqsf). Comme son entraînement s’appuie sur une fonction de récompense basée sur des probabilités plutôt que sur une compréhension logique ou contextuelle du monde, l’IA générative tend à « apprendre une distribution » de probabilités et à tendre vers la médiane de cette distribution, ce qui trop souvent signifie « refléter les pires impulsions, les pires préjugés » qui existent dans leurs données d’apprentissage (a9a3ir). En ce sens, la production de fausses informations échappe largement aux « mauvais acteurs » et devient un phénomène purement technique. Et tandis que les commentateurs réduisent souvent ces problèmes à « l’état actuel de développement » des LLM et affirment qu’ils s’amélioreront avec les progrès techniques (Zagni et Canetta, 2023), il est possible que l’épistémologie même qui sous-tend ces approches les destine à la conjecture statistique plutôt qu’à la production de connaissances faisant autorité.

Compte tenu de cette réalité technique, les personnes rencontrées associent souvent cet enjeu à la problématique générale d’un public mal informé et à la manière dont le battage médiatique sur l’IA donne des impressions irréalistes sur la précision et les compétences de l’IA générative. « Le problème c’est qu’il faut que les jeunes gens puissent se rendre compte que les hallucinations existent et que c’est un résultat prévisible des demandes, explique un interlocuteur, donc le problème ce n’est pas des hallucinations, c’est les gens qui prennent des hallucinations pour du cash » (kzxwsf). Le battage médiatique et la littératie en matière d’IA sont tous deux liés aux risques associés à la désinformation en matière d’IA et peuvent les exacerber. Ici encore, il est essentiel de reconnaître la technicité des hallucinations qui se produisent pour dépasser les limites du cadre des « mauvais acteurs ». Par exemple, les dommages causés par la

désinformation produite par IA peuvent survenir en l'absence de toute mauvaise intention ou même d'acteurs/bénéficiaires humains. Il s'agit de prendre l'exemple d'un spécialiste de l'IA en santé qui s'inquiète de la dépendance excessive à l'égard des résultats des agents conversationnels :

> Obviously this could already happen on the internet that you could just go look up something and get bad information, but it's even worse now. And so that means that, you know, in the worst case scenario we're talking about somebody trying to look up how to do something that's life or death, right? If somebody's like, "Oh no, this person's having an allergic reaction to something, what do I do?" and ChatGPT says some absolute garbage. A person could die, right? Like, that's a significant problem (a9a3ir).

Il ne s'agit là que d'un exemple de la façon dont les choses peuvent mal tourner, même si les chercheurs de Google qui construisent des agents conversationnels spécialisés dans les soins de santé les ont déjà identifiés comme problématiques (Kak & Myers, 2023 ; Singhal & coll., 2023). Alors que les industries et les gouvernements discutent des mesures de sécurité et de garde-fous, il est impossible de prévoir tous les scénarios risqués et problématiques possibles. En l'occurrence, certains des principaux risques découlent directement de la prolifération de « petites » inexactitudes ou de faussetés.

En l'absence d'une véritable littératie sur l'IA, l'utilisation dispersée de ces systèmes et l'acceptation non critique de leurs résultats pourraient avoir des conséquences négatives rapides. Comme l'explique une personne experte :

> ces erreurs d'inférence, ces « hallucinations », par leur utilisation sans esprit critique, deviennent progressivement des faits par leur utilisation par des êtres humains… [et] par les derniers sondages, une grande majorité des personnes qui utilisent ces modèles n'ont pas forcément communiqué à leur employeur ou à leur consommateur [que] le contenu s'est passé par une intelligence artificielle. On se retrouve avec *une sorte de percolation de ces erreurs d'inférence* (zfr5j8).

En d'autres termes, l'utilisation grandissante de l'IA générative dans différentes organisations et institutions de même que par les utilisateurs dans leur vie privée et leur routine quotidienne pose le risque qu'une multiplicité d'inexactitudes soient acceptées et

utilisées sans esprit critique sur le lieu de travail, dans les contextes éducatifs ou dans les communications interpersonnelles, « percolant » ainsi dans toute la société et prenant une aura de vérité à travers son utilisation – souvent sans savoir que l'information provient d'un système d'IA générative ou d'un autre système. « People are already now, unfortunately, using it without understanding what they're seeing as output. Without knowing how this output is generated, they're using it as, you know, information », souligne ainsi une personne (u02uz1). Ce type de problème technique, lorsqu'il est rencontré à grande échelle par le public, pourrait avoir des effets délétères étendus. Au niveau sociétal, s'inquiète une personne interviewée, « we're going to be very, very challenged, I think, to be able to identify reputable sources of information or reliable information » (es6a27). Si les *fake news* et les *deepfakes* compliquent déjà la recherche des faits et de la vérité pour venir perturber la vie politique démocratique, la multiplication des micro-inexactitudes risque d'aggraver le problème d'une manière qu'il est de plus en plus difficile de détecter ou de corriger.

C'est pour cette raison que certaines personnes rencontrées critiquent le déploiement actuel de l'IA, qui est essentiellement dérégulé au Canada, comme s'apparentant plutôt à une « mise en laboratoire » de la société. Dans l'ensemble, l'intégration des agents conversationnels d'IA générative dans les flux de travail des organisations et des appareils gouvernementaux n'a pas suscité de controverse dans le discours public, peut-être hormis quelques inquiétudes dans les contextes journalistiques et éducatifs. L'ampleur de cette intégration de l'IA dans la vie quotidienne des individus augmente ainsi rapidement, mais les problèmes techniques sous-jacents de l'IA générative et les risques qu'ils posent pour la crédibilité de l'information sont rarement, voire jamais débattus publiquement ou par les décideurs politiques. Pour ces raisons, une personne experte s'oppose à la diffusion du ChatGPT auprès d'un public non averti, en suggérant qu'il était

> unethical for them [OpenAI] to use the public for testing their product. Basically, that's what they're doing. And they don't realize the impact it has because people are using it. Like, you know, there are students using it for their courses, there are lawyers that are using it for creating their arguments. There are doctors that, I mean, there are actual doctors that are taking transcribed notes… putting it into the ChatGPT model–and this is all very private health information and they're just putting it out there, you know, not even understanding the implications of that. So in that sense, yeah, it's very irresponsible the way that it was done. The public doesn't have enough awareness or knowledge about these things to be able to use them properly (u02uz1).

Ce qui ressort des entrevues résonne ainsi fortement, et ce surtout à le mettre en échos au débat hautement médiatisé sur les « menaces existentielles ». Alors qu'une grande attention est accordée à des craintes abstraites concernant l'avenir de l'IA, un ensemble de questions plus concrètes concernant quant à elle la désinformation au quotidien soulève des inquiétudes au sein de la communauté canadienne des chercheur.es en IA rencontré.es. Cela vient éclairer et problématiser davantage le domaine complexe de la gouvernance de l'IA. Par exemple, si les menaces les plus immédiates et les plus concrètes de l'IA ne sont pas discutées avec beaucoup de vigueur, alors sur quoi vont s'exercer les politiques et réglementations en matière d'IA ? Comme il s'agira d'en discuter dans quelques instants, l'arène complexe de la gouvernance de l'IA est, pour les personnes rencontrées, également empêtrées dans les lacunes en matière de connaissances sur l'IA en sorte qu'il s'agirait de traduire un ensemble plus diversifié de voix et de connaissances aux discours publics *et* scientifiques qui font l'objet d'un battage médiatique.

4.3 La gouvernance de l'IA se perd-elle dans sa propre traduction ? Tensions démocratiques et expertes dans le façonnement de l'IA

Les enjeux découlant du battage médiatique sur l'IA – et les malentendus réels ou spéculatifs créés – ont des implications importantes pour le devenir de la gouvernance de l'IA. Qui devrait être impliqué dans l'élaboration de la technologie, de la conception à la mise en place de règlements ? Comment atténuer les risques, protéger les citoyens et veiller à ce que les préjudices puissent être reconnus de manière transparente et traités de manière socialement juste et démocratique ? D'une part, les gouvernements semblent confrontés à certaines perspectives peu ou mal informés sur l'IA, ce qui les rend vulnérables à l'influence de différents lobbyistes. D'autre part, les spécialistes en science computationnelle et autres informaticiens n'ont le souvent pas les connaissances sociopolitiques et socioculturelles nécessaires pour comprendre les impacts de l'IA en dehors du laboratoire.

Ce qui est discuté ici présente les points de vue des spécialistes rencontrés sur *qui sont les personnes et quels sont les éléments nécessaires* à l'élaboration de systèmes d'IA et de cadres de gouvernance éthiques et efficaces. Les personnes interviewées reconnaissent massivement la nécessité de rassembler des points de vue interdisciplinaires et d'encourager les collaborations d'experts en sciences sociales, en sciences humaines et d'autres domaines afin de combler les lacunes des connaissances actuelles.

Ces activités sont *intrinsèquement difficiles* à réaliser comme l'indiquent plusieurs expert. es, mais elles sont de la plus haute importance pour développer les langages, valeurs et visions jugés nécessaires à l'élaboration d'une approche éthique et démocratique de l'IA. Impliquer un ensemble plus diversifié de voix et de connaissances et s'affairer à traduire ces connaissances pourrait constituer pour plusieurs un contrepoids et un correctif efficaces – même s'ils sont encore une fois complexes et très débattus – aux questions du battage médiatique sur l'IA, des publics non informés et des attentes irréalistes concernant les capacités de l'IA.

4.3.1 Quelles sont les voix compétentes ? La nécessité d'apports divers dans l'élaboration de l'IA

La plupart des personnes interrogées souhaiteraient encourager l'élaboration d'un cadre réglementaire clair en matière d'IA. La manière dont ce cadre devrait être façonné est toutefois elle ambiguë et controversée, ceci bien qu'un certain nombre de thèmes communs aient été abordés. Par exemple, une approche réglementaire à multiples facettes ou « fragmentée » serait probablement plus appropriée pour certains – par opposition à l'approche plus centralisée du gouvernement fédéral envisagée dans le projet de loi C-27 sur l'intelligence artificielle et les données (LIAD)[40]. Pour d'autres spécialistes de la gouvernance de l'IA, l'approche du gouvernement canadien en la matière n'est pas encore suffisamment solide. Une personne rencontrée soutient qu'Innovation, Sciences et Développement économique Canada (ISDE)) « is not the right place » (qnkqsf) pour rédiger ou appliquer ce type de législation, puisque concentrée principalement sur la commercialisation. En fait, les personnes rencontrées ont exprimé des inquiétudes quant à tout ce qui touche à la *précipitation* lorsque vient le temps de commercialiser et d'implémenter ces technologies – y compris par les gouvernements canadiens. Une personne soutient par exemple qu'il y a « a lot of work to be done » (es6a27) pour aller au-delà de l'approche simpliste du projet de loi, notamment en matière de protection de la vie privée des consommateurs et/ou de son manque de clarté dans les dispositions relatives à l'utilisation « socialement bénéfique » de l'IA. En effet, ces spécialistes font plutôt référence à d'autres organismes gouvernementaux tels que le Bureau de la concurrence du Canada ou le Commissaire à la protection de la vie privée du Canada et à des cadres juridiques qu'ils considèrent comme étant compétents pour réglementer divers éléments plus vastes de l'IA. Une personne considère en outre que

[40] Pour plus de détails sur la LIAD, voir le document d'accompagnement du Gouvernement du Canada à l'adresse suivante : https://ised-isde.canada.ca/site/innover-meilleur-canada/fr/loi-lintelligence-artificielle-donnees-liad-document-complementaire

les efforts très centralisés et vagues du Canada dans le cadre de la LIAD constituent « une approche étrange » qui ne tient pas compte du travail déjà effectué par divers organismes de réglementation et qui apparaît de ce fait comme « maladroite et antidémocratique » (es6a27). Avec les pressions exercées par l'industrie et les Parrains de l'IA tels que Bengio et Hinton en faveur d'une adoption rapide de la LIAD, plusieurs personnes rencontrées trouvent la situation actuelle préoccupante (Deschamps, 2023 ; Panetta, 2023). Une discussion plus soutenue et une approche plus solide – même si elle est plus lente – sont aussi privilégiées et considérées par plusieurs comme étant plus prudentes et respectueuses du processus démocratique.

L'une des principales raisons de s'inquiéter de l'adoption précipitée de réglementations en matière d'IA est, une fois de plus, le manque de connaissances et de compétences apparaissant comme étant généralisées en matière d'IA. Les gouvernements, à savoir autant les décideurs politiques que les administrateurs bureaucratiques, sont perçus comme manquant d'une compréhension technique éclairée de ce qu'est et peut faire l'IA. « I think there's just no understanding [of] it more than a very superficial level in government, of many of the issues of technology, not just [AI] », explique une personne rencontrée (3jpw66). Toujours selon cette personne : « it's easy for somebody to come in from industry and say, 'Here's the truth', and [the government has] no capacity to push back » (3jpw66). Comme le déclare une personne : « les législateurs ne connaissent rien et donc ils sont propices à se faire rouler dans la farine » (kxzwsf). Cette conviction est à la base des critiques formulées à l'encontre de l'ingérence des lobbyistes de la technologie dans l'élaboration des politiques en matière d'IA. « I think the past few years have taught us that there are real limits to the extent that those who are invested in the further development of AI systems can critique their own industry and address some of those [AI] harms », explique une autre personne experte (qnkqsf). Les manques de connaissances techniques exigent donc que pour élaborer des réglementations efficaces et socialement justes en matière d'IA, les agents gouvernementaux prennent le temps d'apprendre des experts techniques non *intéressés* tout en répondant aux besoins du public. Une telle collaboration n'est toutefois pas toujours simple dans la pratique.

Entreprendre un travail interdisciplinaire et traduire les connaissances d'experts dans différents domaines est un défi notoire. Si les décideurs politiques peuvent être confrontés à des lacunes dans les connaissances en matière d'IA qui nécessitent l'intervention d'experts pour éviter l'influence négative des lobbyistes, l'inverse est également vrai. Plusieurs experts en IA ne se sentent pas qualifiés pour commenter

ce qu'ils pensent être le régime réglementaire, car ils n'ont pas les connaissances sociopolitiques approfondies nécessaires pour situer leur travail dans la réalité du monde politique. Comme l'explique une personne rencontrée, « si on parle avec des politiciens, on va parler d'algorithmes, de codes, de choses qui nous intéressent. Mais on ne sera pas dans la vision d'ensemble. […] On n'a pas ces connaissances d'enjeux politiques. On n'a pas ces connaissances d'enjeux éthiques. Zéro » (lwillq). Pour les experts en STEM, ce manque de connaissances sociopolitiques approfondies peut les empêcher de traduire efficacement leur expertise technique pour qu'elle soit pertinente ou compréhensible pour les décideurs politiques.

Selon plusieurs personnes interrogées, les informaticiens et autres spécialistes des STEM ont généralement une vision plutôt myope du monde social et politique, et pourraient de ce fait avoir besoin d'apports et de perspectives plus larges pour façonner un développement et une réglementation éthiques de l'IA. Pour certaines personnes, le problème commence avec la discipline elle-même, à savoir que les informaticiens « are not used to doing cross-disciplinary or interdisciplinary research and [are] not very used to taking seriously expertise from outside » (c3l8ej). Cela s'explique par le fait que l'enseignement de ces disciplines est devenu très spécialisé et rationalisé, faisant fi de la nécessité de suivre des cours de sciences humaines, d'arts ou de sciences sociales. Dans un point de vue autoproclamé « provocateur », une personne affirme qu'une des caractéristiques du milieu

> c'est l'absence de culture ; on a des gens qui n'ont pas de culture sociologique, qui n'ont pas de culture du monde. C'est souvent des gens très pointus dans leur domaine qui sont très forts, mais qui n'ont pas de pratique de l'inférence qui ne se repose pas sur un socle culturel de connaissance des sociétés, de connaissance du monde, etc. (gk3nxq).

Comme le résume un autre expert, « so, if that's the way you're educating people, they're not going to be aware of the context in which their work might be read » (c3l8ej). Peut-être parce qu'ils se concentrent uniquement sur l'amélioration de la technique et non sur les effets secondaires ou distants, les professionnels hautement qualifiés en STEM n'ont pas nécessairement les connaissances pour comprendre l'impact de leurs outils sur le monde. Cette surspécialisation est ainsi considérée comme contribuant au battage médiatique ainsi qu'à des attentes irréalistes, et ce donc, tout en compliquant les efforts de traduction des connaissances techniques aux agents gouvernementaux.

Un ensemble large et interdisciplinaire de perspectives est souhaité pour développer des langages, des valeurs et des visions communes à même de réglementer efficacement et en toute sécurité les systèmes d'IA. « Sociologists should be front and center on teams of academics informing regulators and legislators, » affirme une personne avant de poursuivre en soulignant que « that happens a little bit in Europe, [but] doesn't happen too much in the Canadian jurisdiction because our politicians seem to think that the bureaucrats are sufficiently well educated to be able to draft legislation » (inyw7l). Selon elle, le fait de ne pas inclure les chercheurs en sciences sociales dans le processus d'élaboration des politiques technologiques a pour conséquence que « sometimes they get lucky; most times they introduce consequences they didn't anticipate » (inyw7l). Une telle inclusion ou collaboration est tout aussi importante dans les phases de conception des systèmes d'IA que dans les efforts de réglementation. « Je pense que [les spécialistes des sciences sociales] ont un rôle très important à jouer », explique un interlocuteur (u02uz1). « It's definitely beneficial to have more funding for non-CS, non-engineering fields to engage in AI and machine learning development because they have perspectives that computer scientists and engineers often don't have at their forefront », poursuit-il (u02uz1). Les spécialistes soulignent que la recherche en anthropologie, en communication, en histoire et en sociologie apporte des connaissances « incroyablement importantes » à partager avec les régulateurs, les experts en STEM et les étudiants (u02uz1). Pour beaucoup, ces connaissances sont à la fois source d'espoir et de tension, car elles peuvent constituer un contrepoids à une certaine naïveté, sinon à l'orgueil du domaine des STEM. En fin de compte, elles pourraient également conduire à une adoption de cadres réglementaires beaucoup plus efficaces en mettant un terme aux attentes irréalistes ou au battage médiatique et en veillant à ce que les ingénieurs développent une culture davantage éthique.

Là encore, malgré les appels généralisés à l'interdisciplinarité, les experts soulignent les difficultés à mettre en œuvre ces mêmes collaborations. Alors que certains soulignent l'incompréhension des experts en sciences computationnelles quant à « l'apport que les sciences sociales pouvaient amener » et la tendance à considérer que les chercheurs en sciences sociales ne font qu'aider à obtenir une « social license » pour les technologies, d'autres soulignent les difficultés d'application des connaissances entre les disciplines et la nécessité de « susciter » la collaboration par le biais de programmes de financement spécifiques. Cependant, même avec de tels programmes, ce travail peut s'avérer extrêmement difficile et exigeant. Comme l'explique longuement une personne rencontrée, il existe un besoin d'engagement et de collaboration soutenus plutôt que d'une inclusion superficielle des éthiciens :

> C'est le propre du développement de la recherche intersectorielle : on travaille ensemble au quotidien, on se parle, on est assis côte à côte, c'est en étant côte à côte et en expliquant à l'autre ce que tu fais – alors tu vois là je vais faire ça, les maths derrière, on n'en a rien à carrer, c'est pas le problème en revanche, je vais devoir apprendre tels et tels fondamentaux, il faut que je prenne des sources, comment toi tu vois les choses est-ce que ça c'est légitime, quels sont les sources d'information fiable, quels sont celles sur lesquelles on va pouvoir reposer l'algo (gk3nqx).

Bien que conceptuellement flou, un tel scénario contraste fortement avec les discours dominants sur l'IA éthique ou socialement bénéfique qui traitent le sujet comme une simple dichotomie manichéenne entre les mauvais acteurs/usages et les acteurs bons ou socialement vertueux. Comme l'explique une personne experte, une approche éthique de l'IA implique une adhésion totale à ce qui semble s'apparenter à l'*esprit de confrontation* de la délibération et de la prise de décision démocratiques :

> C'est pas l'algorithme qui est éthique ou pas éthique, c'est sa mise en application, et depuis sa naissance, qui en amont même de sa conception et même en amont la formulation de la question de recherche – ça n'est pas possible que la question de recherche ne soit pas formulée dans une discussion et j'ai presque envie de dire dans une confrontation parce que souvent, c'est confrontationnel, mais au bon sens du terme – il faut que les visions se confrontent, que tu acceptes de confronter ta vision avec celle du sociologue ou du philosophe qui apporte son, et que tu le fasses comprendre ta vision et que lui te fasse comprendre la sienne (gk3nqx).

Ce type d'explication agonistique met en évidence le temps et les efforts collectifs requis afin de susciter des collaborations. Les appels des dirigeants de l'industrie et des Parrains de l'IA à faire adopter à la hâte des réglementations sur l'IA ne tiennent pas ou peu compte du travail long et fastidieux de traduction des connaissances et de valeurs communes sur lesquelles fonder une recherche et une législation rigoureuses, résilientes et socialement justes en matière d'IA. Et, comme si cela n'était pas déjà difficile, cette approche axée sur la vitesse d'adoption ignore le rôle important du demos – le peuple, les citoyens. Comment les citoyens ordinaires, même s'ils ne sont pas parfaitement informés, peuvent-ils contribuer à la conception et à la réglementation de l'IA ? Les personnes rencontrées sont partagées sur la question.

4.3.2 Qu'en est-il de la voix du public ?

La question de savoir si le public doit « avoir voix au chapitre » dans la conception et la gestion des technologies, ou plutôt « [jusqu']à quel point la participation à la prise de décision technique doit-elle s'étendre », a une longue histoire dans les études scientifiques et technologiques (Collins et Evans, 2002, p. 237 ; Marres et coll., 2024). Comme vu auparavant, les personnes interrogées soutiennent généralement l'idée d'une collaboration interdisciplinaire entre spécialistes et la nécessité d'un dialogue avec les agents gouvernementaux. Toutefois, la question de savoir jusqu'où étendre la participation est quelque peu controversée, ce qui témoigne d'une tension importante au sein de l'espace « démocratique » de l'IA canadienne. Quelques experts ne voient par exemple aucune raison d'inclure le public :

> Est-ce qu'il devrait y avoir un débat public ? Je ne vois pas pourquoi moi, j'ai pas été consulté par le clonage humain. Puis tant mieux, parce que Dieu sait que je ne connais rien au clonage et à la biologie. Moi, je pense que ça doit être un débat de spécialistes qui dictent des politiques de politiciens sérieux. Ça c'est déjà difficile. […] c'est que tu peux peut-être faire une démocratisation des savoirs, mais ça prend du temps. Et puis, c'est pour ça, je ne pense pas qu'on puisse avoir un débat public. Je pense qu'on peut expliquer aux gens les règles qui seront mises en place par les gouvernements (lwillq).

Parvenant à une conclusion similaire, un autre expert explique qu'il n'est « pas sûr que la discussion publique soit particulièrement nécessaire », en raison notamment de l'influence démesurée des investisseurs capitalistes et du fait que le public est « extrêmement démuni » par « un différentiel de pouvoir au niveau de la connaissance » (kxzwsf). Pour certains, le niveau général de littératie du public en matière d'IA constitue assez simplement un obstacle à des interventions plus spécialisées dans le débat. De ce point de vue, la solution n'est peut-être pas la participation du public en tant que telle, mais plutôt l'investissement dans « l'alphabétisation » du public.

Comme le rappelle une personne interrogée, il importe de ne pas perdre de vue le potentiel civique et éducatif des appels à faire participer le public à des discussions informatives sur l'IA : « Donc si le développement de l'IA pouvait être l'occasion d'augmenter le niveau général de la population en sensibilité aux problèmes éthiques, ce serait un sacré

gain » (tt3uum). Cette dernière citation est particulièrement frappante, car elle semble résumer un élément clé de l'opinion de nombreux experts. Plusieurs, en effet, ont parlé avec une certaine passion de la nécessité de renforcer la culture de l'IA et la pensée critique dans l'ensemble de la société, considérant souvent qu'il s'agit de la solution la plus directe aux problèmes causés par le battage médiatique sur l'IA. Ces personnes affirment que les citoyens doivent « être plus sceptiques en général » via « une augmentation sans relâche de l'esprit critique ». Néanmoins, les campagnes d'éducation des médias n'ont qu'une portée limitée ; pour garantir que les valeurs démocratiques et les principes de justice influencent la conception et le déploiement de l'IA, il pourrait être nécessaire pour cartain.es d'adopter des approches plus engagées.

Pour d'autres personnes rencontrées, cette ligne de questionnement comporte plusieurs incertitudes. Comme l'explique l'une d'entre elles, « this is theoretically a democracy and therefore the public should choose, with inputs from experts ». Cependant, si l'on considère l'IA d'un point de vue plus technique, une hésitation se fait sentir :

> But there's arguments against it as well. Should, if one takes sort of a purely tools-based view of AI, like we don't ask the public when we buy a new x-ray machine or when we adopt a new standard of care, like a new workflow or something, [so] why should we ask them in this specific instance? I am sympathetic to both. I don't have an opinion yet, I guess. I'm actively working through this myself, so I don't actually know what the right answer is (p9efse).

Cette hésitation est – au moins implicitement – commune à plusieurs personnes qui estiment qu'ils ne sont pas qualifiés pour parler des questions de gouvernance au-delà de leur opinion en tant que citoyens. Aussi, malgré les nombreuses frictions en jeu, plusieurs personnes interviewées sont convaincues que les citoyens et les communautés de toutes sortes doivent être entendus, beaucoup d'entre eux appelant à la création d'une sorte de « consortium » qui inclurait un vaste éventail d'intervenant.es, des grandes entreprises aux « petites et moyennes entreprises, universités, gouvernements, hôpitaux et membres de la communauté, y compris les organisations à but non lucratif et les organisations de défense des droits numériques » (u54r89). Plus important encore, une personne experte souligne l'importance d'inclure les voix les plus éloignées des structures de pouvoir existantes :

> the other voices [needed] are the people that are actually affected by these systems. For a long time, as these systems were sort of germinating within the industry and academia, they were easy to ignore. But you know, now these systems are deployed at scale, they affect millions of people. There's a reason why the most important and vocal critiques of these systems came from Black women, from trans folks, from people on the margins who experience a lot of these harms of these systems first and most acutely. There needs to be a way to take those voices seriously and to integrate the critiques and the harms that they're identifying into the way we deal with these systems (qnkqsf).

En outre, les spécialistes des études noires et des études critiques sur la race ont retracé la longue histoire et l'héritage contemporain du racisme médiatisé par la technologie, qui s'est infiltré dans la conception même – sans parler de l'application sociale – des technologies numériques contemporaines (Benjamin, 2019 ; Browne, 2015 ; Atanasoski et Vora, 2019 ; Wang, 2018). Il est ainsi important de veiller à ce que les membres les plus marginalisés de la société aient la possibilité d'exprimer leurs expériences vécues et leurs préoccupations concernant la croissance et le déploiement de l'IA de manière à en façonner la trajectoire et la réglementation.

Les nombreux problèmes engendrés par le déploiement social de l'IA sont, dans l'ensemble, articulés dans des domaines ou des contextes *spécifiques* – les laboratoires d'informatique, les couloirs des décideurs politiques, les départements de sciences sociales et humaines, les entreprises privées, etc. Pour plusieurs personnes interrogées, le défi consiste à communiquer et à traduire efficacement ces problématiques *dans tous les domaines* afin que les discussions sur les risques et les solutions puissent être articulées et évaluées avec plus de nuances et de profondeur contextuelle. Cette question de l'application des connaissances concerne toutes les parties prenantes et exige un effort de coopération entre les scientifiques et les universitaires, les développeurs d'IA, les entreprises, les journalistes, les décideurs politiques et les personnalités médiatisées telles que les Parrains de l'IA, dont aucun ne devrait avoir le monopole de la discussion. Les personnes rencontrées s'entendent toutes sur ce point : afin de surmonter les lacunes en matière de connaissances et les perspectives partielles sur l'IA – à chaque étape, de la conception à la gouvernance –, les efforts de recherche collaborative et d'application des connaissances seront essentiels. L'enjeu ou la problématique en est ainsi une d'application de ce consensus au moins théorique. Y arriver pourrait contribuer grandement à contrer les préjudices sociaux et les risques

pour le domaine de l'IA évoqués plus haut, en partie en modifiant les imaginaires et les attentes irréalistes – non seulement parmi les publics, mais aussi parmi les développeurs experts de l'IA eux-mêmes.

CONCLUSION

Ce rapport est à la fois large et précis. Large, il l'est dans la mesure où il fait partie d'un vaste projet comparatif entre différents « niveaux » ou layers de la construction sociale de l'IA allant des médias, à la gouvernance ainsi qu'à la société civile de même qu'entre les quatre pays du consortium *Shaping AI* que sont la France, le Royaume-Uni, l'Allemagne et le Canada. Précis, il l'est dans son attachement à patiemment recueillir les avis et analyses des personnes situées en amont du déploiement de l'IA au Canada, à savoir les chercheur.es qui œuvrent dans les laboratoires, les instituts et autres officines universitaires. Fait rare dans la littérature, l'équipe canadienne de *Shaping AI* est allée sur le *terrain de la recherche* en train de se faire au pays, c'est-à-dire qu'elle est allée à la rencontre de pratiques et des positionnalités qui passent souvent « sous le radar », pour ainsi dire, aussi loin qu'elles puissent être de l'attention médiatique ou des gouvernements[41]. Ces pratiques et positions ne peuvent être neutres par définition ; aussi, ce que l'équipe de recherche a tenté de comprendre, c'est la signification que les personnes interviewées leur accordaient : les valeurs, attentes et justifications motivant ce qui se fait au quotidien de la vie de laboratoire, mais également le sens plus vaste qui peut être octroyé à la place du développement de l'IA en société. Ce qui a ainsi été découvert est bien toute l'adaptabilité et ce que Pinch et Bijker nomment la « flexibilité interprétative » des propos tenus. En l'occurrence, même l'opposition élémentaire entre, d'une part, les chercheur.es plus « computo-centriques » et ceux et celles, d'autre part, davantage « socio-centriques » n'est pas aussi nette dans les faits. Il y a plutôt continuum ou arc avec des inclinaisons certes senties, mais aussi le plus souvent – et de manière sans doute étonnante – des appels au dialogue entre les différentes expertises qui sont eux aussi plutôt sentis.

Une des principales clés pouvant expliquer cette adaptabilité des chercheur.es rencontré.es tient à la nature même de l'objet auquel ils s'attachent. L'IA est une science ambiguë, dissonante et multivoque. Ce qui est apparu au cours des entrevues sont

[41] Pour rappel méthodologique, les entrevues semi-dirigées qui ont été effectuées l'ont été avec des personnes faisant de l'IA, mais en dehors des centres plus usuels que sont le MILA, le Vector et l'AMII et ce, entre autres parce que les chercheurs de ces institutions se sont montrés assez peu enclin à dialoguer avec les chercheurs en sciences humaines et sociales tels ceux de l'équipe Shaping AI.

ainsi les traces de différents tâtonnements et d'efforts afin de rendre intelligible un déploiement scientifique et technologique emplies d'incertitudes. Pour tautologique que cela puisse paraître, l'IA est une construction sociale qui renvoie à une *négociation sociale* ; dite négociation à laquelle participent les spécialistes rencontré.es. Ce que le rapport montre à cet égard tient à une forme pour le moins particulière de débat. Quelles sont les grandes controverses ? Qui s'opposent et pourquoi ? La réponse à ces questions est elle aussi ambiguë. Sur le terrain apparaissent des divergences de points de vue, de perspectives qui sont autant de traductions de ce qui est présagé par l'IA, mais, somme toute, assez peu de réels et profonds conflits d'interprétation. Cela a par ailleurs tendance à donner raison à Lucy Suchmann lorsqu'elle parle aujourd'hui d'une certaine « uncontroversial thingness » de l'intelligence artificielle. Au Canada, il n'y a pas eu et il n'y a pas présentement de grand scandale ou de grande « affaire » – au sens tout français du terme – autour desquels puisse s'enrouler l'attention. La, ou plutôt les tensions et lignes de fuite sont plus subtiles et, justement, plus ambiguës.

Le premier chapitre du rapport s'intéresse précisément à cet enjeu d'une définition *en passe de se faire* de l'IA. Les personnes rencontrées s'interrogent ouvertement sur quelque chose qui a tout d'un problème évanescent : celui de l'ambiguïté même de ce qu'est et fait l'intelligence artificielle. Cette réalité – celle entre autres du *modus operandi* des LLM – l'inscrit dans une forme incertaine de débat quant à ce que peuvent vouloir dire des notions telles que celles de compréhension, d'explicabilité ou d'interprétabilité. Chaque fois, c'est la taille et l'opacité des modèles qui posent problème : qu'est-ce qui est tiré et retiré de leurs inférences ; qu'est-ce qui lie logiquement les différentes étapes d'un fonctionnement x ou y ; quel en est plus largement le sens véhiculé ou encore sa portée en société ? Aussi, de ce type d'enjeu à des questions de nature plus communicationnelle touchant entre autres aux différentes disciplines interpellées par l'IA, le pas est relativement facile à franchir. Si l'histoire de l'IA n'est pas la chose la plus connue et s'il existe encore et toujours de nombreux glissements terminologiques, c'est bien le dialogue intradisciplinaire souhaité par les uns et les autres qui se montre difficile. En particulier, il importe de ne pas sous-estimer les dissonances qui peuvent exister avec les sciences humaines et sociales alors que ces dernières ne peuvent être réduites à un rôle de cautionnement, pour des raisons logiques et pratiques. En somme, ce premier chapitre est ce qui permet d'explorer la *cooccurrence* de problèmes affectant la définition même de l'IA, tant au niveau de la constitution des systèmes technologiques, que dans leur traitement de la signification – la sienne propre et celle du monde extérieur – pour se répercuter *in fine* dans les échanges entre les acteurs du monde scientifique.

Le second chapitre s'intéresse également à une controversialité *de fond*, c'est-dire à un enjeu difficile à cerner à la surface des discours promotionnels de l'IA ou à travers son traitement dans les médias. Au détour des entrevues avec les spécialistes, en effet, c'est la matérialité et la technicité de l'IA comme nouvelle forme de calcul qui est apparue. Augmenter la taille des corpus, le nombre de paramètres des modèles, la vitesse de calcul des algorithmes, etc., tout ceci n'est pas qu'un mode ou une possibilité, mais plutôt un type d'obligation au sein de la concurrence que se livrent les chercheur.es en IA et leurs commanditaires. L'*upscaling* est en ce sens un *régime*, à savoir une réalité à la fois structurée et structurante pour ce secteur de la recherche. Ce qui apparait alors sur le terrain canadien de la recherche sont des conditions qui appartiennent à une économie politique particulière. Le Canada est un écosystème somme toute centralisé dans lequel les promesses technoscientifiques – hyperboliques, maisx aussi parfois dystopiques – donnent accès aux deniers, données et infrastructures de calcul. Ici encore, pour surprenant que cela puisse paraître à première vue, plusieurs des chercheur.es interviewé.es se montrent critiques des derniers développements, notamment dans leur actualisation par le biais d'une problématique écologique. Ce qui est souvent décrit, aussi, est un phénomène de débordement ou de fuite en avant à travers lesquels l'écosystème canadien perd en contrôle, à savoir, surtout en souveraineté face à une science de plus en plus industrialisée et de plus en plus dictée à partir d'impératifs d'origine américaine.

Pour sa part, le troisième et dernier chapitre fait état des réponses reçues en ce qui a trait au devenir social de l'IA, à savoir ce qu'elle peut bien faire une fois déployée au-delà des laboratoires. Le chapitre débute en parlant d'*hype* et des discours promotionnels faisant les louanges des prouesses de la technologie. De manière quelque peu surprenante, encore cette fois, les expert.es interrogé.es sont plutôt d'avis que le battage médiatique actuel est devenu tout aussi répandu que problématique. Tout se passe comme si l'engouement était devenu contre-productif ou à même de saper son propre entrain. Pour plusieurs personnes rencontrées, le problème réside dans le fait que les discours hyperboliques finissent le plus souvent par affecter des publics peu ou pas informés à propos des capacités réelles des systèmes d'intelligence artificielle, ceci faisant en sorte qu'ils soient vulnérables à ce type de matraquage publicitaire. Au fil des entrevues, l'idée que l'IA soit considérée comme une solution « quasi magique » pour tout un éventail de choses est revenue à plusieurs reprises. Cela tient entre autres à l'écho provenant de l'espace médiatique, mais, plus encore, il s'agit de voir que ce même écho se réverbère pour venir modifier la qualité même de la science produite. Un exemple de ce

type d'empressement et de dépassement extrascientifique se trouve dans la discussion autour des « risques existentiels » promus par les Parrains de l'IA tels que Joshua Bengio et Geoffrey Hinton. La majorité des personnes rencontrées estiment que ce discours est assez peu argumentatif – du moins, d'un point de vue de sciences humaines et sociales – et que, de ce fait, il vient détourner l'attention de questions plus immédiates et ancrées socialement. À ce propos, l'un des principaux sujets de préoccupation relevés concerne les risques que l'IA générative fait peser sur l'écosystème de l'information. Un second enjeu ayant une résonance chez les personnes rencontrées relève de la gouvernance de l'IA, c'est-à-dire qui devraient être entendu et qui devraient, pour ainsi dire, avoir son mot à dire en termes de réglementation parmi les experts, qu'ils soient « computo » ou « socio »-centriques, et les citoyens, surtout ceux qui sont les plus touchés par les conséquences négatives du déploiement de l'IA.

La question qui se pose, *in fine*, est sans doute la suivante : pourquoi faut-il donc que la controversialité de l'IA soit si intimement liée à l'enjeu de sa gouvernance ? Et, d'abord, qu'est-ce que la gouvernance de l'IA au-delà des discours les plus usuels ? Son origine étymologique, du verbe grec *kubernân* au verbe latin *gubernare*, indique assez bien de quoi il s'agit : de pilotage, de navigation dans des eaux incertaines et d'orientation dans un futur par définition encore inconnu. Tout ceci qui s'applique au déploiement de l'IA à la sortie des laboratoires et lors de son entrée dans la société. Quel discours contrôle quoi et avec quelle légitimité ? Ce qui a été vu au long de ce rapport, c'est comment une pléthore de spécialistes peut offrir une pléthore de valeurs, d'actions et d'interprétations. Le pluralisme est en ce sens relativement souverain. Le problème est que certaines de ces valeurs valent davantage que d'autres et qu'elles se traduisent mieux que d'autres. Ce qui vaut également pour leur porte-parole. À paraphraser *La ferme des animaux* de George Orwell, il s'agirait de souligner que tous « sont égaux, mais certains sont plus égaux que d'autres » (1945). L'écosystème de l'IA au Canada en est un pour le moins centralisé. Le plus souvent, ce sont les mêmes porte-paroles qui promeuvent les mêmes idées. Le plus souvent, aussi, ce sont eux qui attirent le plus de ressources, que celles-ci soient de nature économique, politique ou médiatique. D'où, justement, l'importance de cette question de la gouvernance. Ce que ce rapport a tenté de démontrer est comment non seulement ce pluralisme des voix existe – si tant est qu'il est recherché auprès des multiples spécialistes sur le terrain – mais encore que ce pluralisme ait certainement quelque chose d'un *desideratum* dans l'état actuel de la recherche en IA au Canada. Avoir nombre de voix au sein *des* sciences et technologies de l'IA possède sa propre vertu, mais encore, elle est ce qui permet de penser une multiplicité de réponses politiques,

médiatiques et, même, d'entrevoir ce qui pourrait être une capacité de la société civile à élaborer sa propre réplique.

BIBLIOGRAPHIE

Akrich, M., Callon, M., Latour, B., & Monaghan, A. (2002). *The key to success in innovation part i : The art of interessement. International Journal of Innovation Management, 06*(02), 187-206. https://doi.org/10.1142/S1363919602000550

Akrich, M. (2006). La description des objets techniques. Dans M. Akrich, M. Callon, & B. Latour (Éds.), *Sociologie de la traduction: textes fondateurs* (p. 159 178). Presses des Mines. https://doi.org/10.4000/books.pressesmines.1197

Amoore, L. (2019). Doubt and the Algorithm : On the Partial Accounts of Machine Learning. *Theory, Culture & Society, 36*(6), 147 169. https://doi.org/10.1177/0263276419851846

Ananny, M., & Crawford, K. (2018). Seeing without knowing : Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973 989. https://doi.org/10.1177/1461444816676645

Anderson, C. (23 juin 2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. Consulté 30 septembre 2024, à l'adresse https://www.wired.com/2008/06/pb-theory/

Atanasoski, N., & Vora, K. (2019). *Surrogate humanity: Race, robots, and the politics of technological futures*. Duke University Press.

Attard-Frost, B. (2022, February 21). Opinion: Once a promising leader, Canada's artificialintelligence strategy is now a fragmented laggard. *The Globe and Mail*. https://www.theglobeandmail.com/opinion/article-once-a-promising-leader-canadas-artificial-intelligence-strategy-is/

Attard-Frost, B., & Widder, D. G. (2023). The Ethics of AI Value Chains: An Approach for Integrating and Expanding AI Ethics Research, Practice, and Governance | Montreal AI Ethics Institute. Retrieved February 23, 2024, from https://montrealethics.ai/the-ethics-of-ai-value-chains-an-approach-for-integrating-and-expanding-ai-ethics-research-practice-and-governance/

Bellon, A., & Velkovska, J. (2023). L'intelligence artificielle dans l'espace public : Du domaine scientifique au problème public: Enquête sur un processus de publicisation controversé. *Réseaux, 240*(4), 31-70. https://doi.org/10.3917/res.240.0031

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Bender, E. M. et Koller, A. (2020). Climbing towards NLU : On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185 5198. https://doi.org/10.18653/v1/2020.acl-main.463

Bengio, Y. (2016). The Rise of Artificial Intelligence through Deep Learning | Yoshua Bengio | TEDxMontreal, *Youtube*. https://youtu.be/uawLjkSI7Mo?si=IaZRhrnKCZYVGHlD&t=17.

Bengio, Y., Cohen, A., Prud'homme, B., DE Lima Alves, L. A. & Oder, N. (2023). Innovation ecosystems for socially beneficial AI. In Prud'homme, B., Régis, C., Farnadi, G., Dreier, V. & Rubel, S. (dir). *Missing links in AI governance*, (1st ed, p.133-148). UNESCO and MILA.

Benjamin, R. (2019) *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity Press.

Bergen, M., wt Wagner, K. (2015, July 15). Welcome to the AI Conspiracy: The "Canadian Mafia" Behind Tech's Latest Craze. *Vox*. https://www.vox.com/2015/7/15/11614684/ai-conspiracy-the-scientists-behind-deep-learning

Birch, K. (2020). Technoscience Rent: Toward a Theory of Rentiership for Technoscientific Capitalism. *Science, Technology, & Human Values, 45*(1), 3-33. https://doi.org/10.1177/0162243919829567

Birhane, A., Kalluri, P., Card, D., Agnew, W., Dotan, R., & Bao, M. (2022). The Values Encoded in Machine Learning Research. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173-184. https://doi.org/10.1145/3531146.3533083

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., & al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology analysis & strategic management, 18*(3-4), 285-298.

Bostrom, N. (2013). Existential risk prevention as global priority. *Global Policy, 4*(1), 15-31.

Brkan, M. (2019). Artificial Intelligence and Democracy: The impact of disinformation, social bots and political targeting. *Delphi*, 2, 66-71. DOI: 10.21552/delphi/2019/2/4

Browne, S. (2015). *Dark Matters: On the surveillance of blackness*. Duke University Press.

Canadian Press. (2020, May 12). Postmedia to Lay Off 80, Permanently Close 15 Newspapers amid COVID-19 Fallout. https://www.bnnbloomberg.ca/postmedia-to-lay-off-80-permanently-close-15-newspapers-amidcovid-19-fallout-1.1428207

Callon, M. (1980). Struggles and Negotiations to Define What is Problematic and What is Not. In K. D. Knorr, R. Krohn, & R. Whitley (Eds.), *The Social Process of Scientific Investigation* (pp. 197–219). Springer Netherlands. https://doi.org/10.1007/978-94-009-9109-5_8

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. *The Sociological Review, 32*(1_suppl), 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Callon, M. (1984). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay. The Sociological Review, 32(1_suppl), 196-233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Callon, M. (1986). Éléments pour une sociologie de la traduction : La domestication des coquilles Saint-Jacques et des marins pêcheurs dans la baie de Saint-Brieuc. *L'Année Sociologique, 36*, 169-208.

Cardon, D. (2015). *À quoi rêvent les algorithmes? Nos vies à l'heure des* big data. Seuil.

Cardon, D., Cointet, J. & Mazières, A. (2018). La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle. *Réseaux, 211*, 173-220. https://doi.org/10.3917/res.211.0173

Center for AI Safety. (2023). Statement on AI Risk: AI experts and public figures express their concern about AI risk. https://www.safe.ai/statement-on-ai-risk

Chartier-Edwards, N., Grenier, E., & Roberge, J. (sous presse). If (world2vec) then vec2politics : On Machine Learning and the Performativity of Recursive Power. In D. Brzezinski, K. Filipek, K. Piwowar, & W.-B. Malgorzata (Éds.), *Algorithms, Artificial intelligence and beyond : Theorising Society and Culture of the 21st Century*. Routledge.

Christin, A. (2020). *Metrics at Work*. Princeton University Press. https://press.princeton.edu/books/ebook/9780691200002/metrics-at-work

Collins, H. M., & Evans, R. (2002). The Third Wave of Science Studies: Studies of Expertise and Experience. *Social Studies of Science, 32*(2), 235-296.

Collins, H. M., & Evans, R. (2007). *Rethinking Expertise*. University of Chicago Press.

Colleret, M., & Gingras, Y. (2022). L'intelligence artificielle au Québec: « révolution » et ressources publiques. In G. Dandurand, F. Lussier-Lejeune, D. Letendre, & M.-J. Meurs (Eds.), *Attentes et promesses technoscientifiques* (pp. 75–96). Presses de l'Université de Montréal.

Compton, J., & Dyer-Witheford, N. (2014). Prolegomenon to a theory of slump media. *Media, Culture & Society, 36*(8), 1196–1206.

Crawford, K. (2021). *The Atlas of AI.* Yale University Press.

Dandurand, G. et al. (2020). Social Dynamics of Expectations and Expertise: AI in Digital Humanitarian Innovation. *Engaging Science, Technology, and Society, 6*, 591-614. DOI:10.17351/ests2020.459.

Dandurand, G., et al. (2022). Training the News: Coverage of Canada's AI Hype Cycle (2012-2021), Shaping 21st-Century AI, Institut national de la recherche scientifique. https://espace.inrs.ca/id/eprint/13149/1/report_ShapingAI_verJ.pdf

Dandurand, G., Blottière, M., Jorandon, G., Gertler, N., Wester, M., Chartier-Edward, N., Roberge, J. & McKelvey, F. (2023). Entraîner l'actualité: La couverture canadienne du cycle d'engouement pour l'IA (2012-2021).

Dandurand, G., McKelvey, F., & Roberge, J. (2023). Freezing out: Legacy media's shaping of AI as a cold controversy. *Big Data & Society, 10*(2). https://doi.org/10.1177/20539517231219242.

Davidson, J. (2023, 30 October). Google Brain founder says big tech is lying about AI extinction danger. *Australia Financial Review*. https://www.afr.com/technology/google-brain-founder-says-big-tech-is-lying-about-ai-human-extinction-danger-20231027-p5efnz

Deschamps, T. (2023, 24 October). AI pioneer Yoshua Bengio wishes Canada's AI legislation was further along by now. *CTV News Montreal*. https://montreal.ctvnews.ca/ai-pioneer-yoshua-bengio-wishes-canada-s-ai-legislation-was-further-along-by-now-1.6615272

Dyer-Witheford, N., Kjosen, A. M., et Steinhoff, J. (2019). *Artificial Intelligence and the Future of Capitalism*. Pluto Press.

Elish, M. C., et boyd, danah. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs, 85*(1), 57 80. https://doi.org/10.1080/03637751.2017.1375130

Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research policy, 29*(2), 109-123.

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. (2023, 30 October). *The White House*. https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Fenton, N. (2011). Deregulation or democracy? New media, news, neoliberalism and the public interest. *Continuum: Journal of Media & Cultural Studies, 25*(1), 63–72.

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication, 37*(2), 145–156. https://doi.org/10.1080/10584609.2020.1723755

Foucault, M. (2001). *Dits et écrits, 1*. Gallimard

Foucault, M. (2012). *Du gouvernement des vivants. Cours au Collège de France 1979–1980*. Seuil/Gallimard.

Forget, P. (2019, 13 septembre). 200M$ de plus pour Element AI. *Les Affaires*. https://www.lesaffaires.com/auteur/pascal-forget/2280

Forsythe, D. E. (1993). Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. *Social Studies of Science, 23*(3), 445–477. https://doi.org/10.1177/030631279302300302

Future of Life Institute. (2023, 22 March). Pause Giant AI Experiments: An Open Letter. Future of Life Institute. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Gebru, T., Bender, E. M., McMillan-Major, A., & Mitchell, M. (2023, 31 March). Statement from the listed authors of Stochastic Parrots on the 'AI pause' letter. *DAIR*. https://www.dair-institute.org/blog/letter-statement-March2023/

George-Cosh, D. (2018, July 1). Element AI aims for unicorn status with record Canadian financing: Sources. *BNN Bloomberg*. https://www.bnnbloomberg.ca/element-ai-aims-for-unicorn-status-with-record-canadian-financing-sources-1.1101206

Gold, A., & Fischer, S. (2023, 21 February). Chatbots trigger next misinformation nightmare. *Axios*. https://www.axios.com/2023/02/21/chatbots-misinformation-nightmare-chatgpt-ai

Gouvernement du Canada, (2022, 20 août). *Stratégie pancanadienne en matière d'intelligence artificielle*. https://ised-isde.canada.ca/site/strategie-ia/fr

Government of Canada. (2023, June 9). *Government of Canada invests in responsible artificial intelligence research at the Université de Montréal*. https://www.canada.ca/en/innovation-science-economic-development/news/2023/06/government-of-canada-invests-in-responsible-artificial-intelligence-research-at-the-universite-de-montreal.html

Gouvernement du Québec (2023, 27 avril). Encadrement de l'intelligence artificielle (IA) - Le gouvernement mandate le Conseil de l'innovation du Québec pour organiser une consultation sur l'IA. https://www.quebec.ca/nouvelles/actualites/details/encadrement-de-lintelligence-artificielle-ia-le-gouvernement-mandate-le-conseil-de-linnovation-du-quebec-pour-organiser-une-consultation-sur-lia-47386

Guice, J. (1999). Designing the future: The culture of new trends in science and technology. *Research Policy, 28*(1), Article 1. https://doi.org/10.1016/S0048-7333(98)00105-X

Halin, F., & Larocque, S. (2020, November 30). Une mine d'or de savoirs cédée aux Américains. *Le Journal de Montréal*. https://www.journaldemontreal.com/2020/11/30/element-ai-avalee-par-une-americaine

Hempel, J. 2017. Inside Microsoft's AI Comeback, *Wired*. June 21, 2017. https://www.wired.com/story/inside-microsofts-ai-comeback/

Hoffman, S. G. (2017). Managing ambiguities at the edge of knowledge: Research strategy and artificial intelligence labs in an era of academic capitalism. *Science, technology, & human values, 42*(4), 703-740.

Hoffmann, A. L. (2021). Terms of inclusion: Data, discourse, violence. *New Media & Society, 23*(12), 3539-3556.

Holton, R., & Boyd, R. (2021). 'Where are the people? What are they doing? Why are they doing it?'(Mindell) Situating artificial intelligence within a socio-technical framework. *Journal of Sociology, 57*(2), 179–195. https://doi.org/10.1177/1440783319873046

Hsu, T., & Thompson, S. A. (2023, 8 February). Disinformation Researchers Raise Alarms About A.I. Chatbots. *The New York Times*. https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html

Hunt, R., & LePage-Richer, T. (2024). AI and the Canadian Institute for Advanced Research. In F. McKelvey, S. Toupin, & J. Roberge (Éds.), *Northern Lights and Silicon Dreams : AI Governance in Canada (2011-2022)* (p. 74-94). Shaping AI.

Innovation, Science and Economic Development Canada. (n.d.). The Artificial Intelligence and Data Act (AIDA) – Companion Document. https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document

Kak, A., & Myers West, S. (2023, 9 November). The AI Debate Is Happening in a Cocoon. *The Atlantic*. https://www.theatlantic.com/ideas/archive/2023/11/focus-problems-artificial-intelligence-causing-today/675941/

Kaplan, A. (2020). Artificial Intelligence, Social Media, and Fake News: Is this the end of democracy? In Gül, A. A., Ertürk, Y. D., & Elmer, P. (Eds.), *Digital Transformation in Media & Society* (pp. 149–162). DOI: 10.26650/B/SS07.2020.013

Kertysova, K. (2018). Artificial Intelligence and Disinformation: How AI changes the way disinformation is produced, disseminated, and can be countered. *Security and Human Rights, 29,* 55–81. doi:10.1163/18750230-02901005

Latour, B. (1987). *Science in action : How to follow scientists and engineers through society*. Harvard University Press.

Latour, B. (2005). *Reassembling the Social : An Introduction to Actor-Network-Theory*. OUP Oxford.

La Presse canadienne. (2019, September 14). Element AI accueille la Caisse parmi ses investisseurs. *Le Devoir*. https://www.ledevoir.com/economie/562608/element-ai-accueille-la-caisse-parmi-ses-investisseurs

Lepage-Richer, T. (2021). Adversariality in Machine Learning Systems: On Neural Networks and the Limits of Knowledge. In J. Roberge & M. Castelle (Eds.), *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies* (pp. 197-225). Springer International Publishing. https://doi.org/10.1007/978-3-030-56286-1_7

Lepage-Richer, T., & McKelvey, F. (2022). States of computing: On government organization and artificial intelligence in Canada. *Big Data & Society, 9* (2), 205395172211233. https://doi.org/10.1177/20539517221123304

Lipton, Z. & Steinhardt, J. 2018. Troubling Trends in Machine Learning Scholarship, ArXIV, July, 26, 2018. https://arxiv.org/abs/1807.03341

Mackenzie, A. (2015). The production of prediction: What does machine learning want?. *European Journal of Cultural Studies, 18*(4-5), 429-445.

Magaudda, P. (2014). The Broken Boundaries between Science and Technology Studies and Cultural Sociology : Introduction to an Interview with Trevor Pinch. *Cultural Sociology, 8*(1), 6376. https://doi.org/10.1177/1749975513484604

Marinov, R. (2020). Mapping the infotainment literature: current trajectories and suggestions for future research. *The Communication Review, 23*(1). https://doi.org/10.1080/10714421.2019.1682894.

Marres, N., & Gerlitz, C. (2016). Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology. *The Sociological Review, 64*(1), 21-46. https://doi.org/10.1111/1467-954X.12314

Marres, N., Castelle, M., Gobbo, B., Poletti, C., & Tripp, J. (2024). AI as super-controversy : Eliciting AI and society controversies with an extended expert community in the UK. *Big Data & Society, 11*(2), 20539517241255103. https://doi.org/10.1177/20539517241255103

Masbourian, P. 2023. Moratoire demandé sur l'IA : « C'est comme l'arrivée du nucléaire », *Tout un matin*, (réalisateur), Radio-Canada, 30 mars 2023. https://ici.radio-canada.ca/ohdio/premiere/emissions/tout-un-matin/segments/entrevue/438367/moratoire-intelligence-artificielle-arrivee-nucleaire

McCarthy, J. et al. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

McKenna, A. (21 septembre 2023). Le Québec ne patente plus comme avant. *Le Devoir*. https://www.ledevoir.com/economie/798493/recherche-et-developpement-quebec-ne-patente-plus-comme-avant

McKelvey, F. (2018). *Internet Daemons: Digital Communications Possessed*. University of Minnesota Press.

McKelvey, F. (2021). The other cambridge analytics: early “artificial intelligence” in American political science. In Roberge, J. & Castelle, M. (eds.). *The Cultural Life of Machine Learning: An Incursion into Critical AI Studies*, 117-142. Palgrave MacMillan.

McKelvey, F. (2023, 3 April). Let’s base AI debates on reality, not extreme fears about the future. *The Conversation*. https://theconversation.com/lets-base-ai-debates-on-reality-not-extreme-fears-about-the-future-203030

McKelvey, F., Toupin, S., & Roberge, J. (Éds.). (2024). *Northern Lights and Silicon Dreams : AI Governance in Canada (2011-2022)*. Shaping AI.

McKinsey. (2023, November 20). The state of AI in 2023: Generative AI’s breakout year. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-AIs-breakout-year

Mendon-Plasek, A. (2020). Mechanized signifiance and machine learning : Why it became thinkable and préférable to teach mchines to judge the world. Dans J. Roberge & M. Castelle (dir). *The cultural life of machine learning : An incursion into critical AI studies.*, 13-78.. Palgrave macmillan.

Merton, R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review, 8*(2), 193-210. https://doi.org/10.2307/4609267

Mol, A. (2002). *The Body Multiple: Ontology in Medical Practice*. Duke University Press.

Morrish, L. (2023, 1 February). Fact-Checkers Are Scrambling to Fight Disinformation With AI." *Wired*. https://www.wired.com/story/fact-checkers-ai-chatgpt-misinformation/

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics*. https://doi.org/10.1007/s43681-022-00209-w

Nakonechny, S. (2024). AI pioneer Yoshua Bengio urges Canada to build $1B public supercomputer. *CBC News*. https://www.cbc.ca/news/canada/montreal/bengio-asks-canada-to-build-ai-supercomputer-1.7094858

OECD. (2023, 7 September). *G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI. Report Prepared for the 2023 Japanese G7 Presidency and the G7 Digital and Tech Working Group*. OECDpublishing. https://www.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence ai_bf3c0c60-en

Onstad, K. (2018, 29 janvier). Mr. Robot. *Toronto Life*. The AI superstars at Google, Facebook, Apple–they all studied under this guy (torontolife.com)

Panetta, A. (2023, 26 July). Canadian AI pioneer brings plea to U.S. Congress: Pass a law now. *CBC News*. https://www.cbc.ca/news/world/ai-laws-canada-us-yoshua-bengio-1.6917793

Pasquinelli, M. & Joler, V. (2020). The nooscope manifested : AI as instrument of knowledge extractivisme. *AI & Society, 36*, p.1263-1280. https://doi.org/10.1007/s00146-020-01097-6

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts : Or How the Sociology of Science and the Sociology of Technology might Benefit Each Other. *Social Studies of Science, 14*(3), 399-441. https://doi.org/10.1177/030631284014003004

Radford, J., & Joseph, K. (2020). Theory In, Theory Out : The Uses of Social Theory in Machine Learning for Social Science. F*rontiers in Big Data, 3*, 18. https://doi.org/10.3389/fdata.2020.00018

Rella, L. (2023). Close to the metal: Towards a material political economy of the epistemology of computation. *Social Studies of Science, 0*(0). https://doi.org/10.1177/03063127231185095

Roberge, J., Morin, K., & Senneville, M.. (2019). Deep Learning's Governmentality. In Sudmann, A. (ed.). *The democratization of artificial intelligence: Net politics in the Era of learning algorithms,* 123-142. Transcript Verlag.

Roberge, J., Senneville, M., & Morin, K. (2020). How to translate artificial intelligence? Myths and justifications in public discourse. *Big Data & Society, 7*(1). https://doi.org/10.1177/2053951720919968

Roberge, J., & Lebrun, T. (2021). BERT, GPT-3, Timnit Gebru et nous: l'intelligence artificielle à la conquête du langage. *Sociologie et sociétés, 53*(1), 235-257.

Roberge, J., & Castelle, M. (Eds.). (2021). *The Cultural Life of Machine Learning*. Palgrave Macmillan. https://www.beck-shop.de/roberge-castelle-cultural-life-of-machine-learning/product/31688490

Roberge, J., Dandurand, G., Morin, K., & Senneville, M. (2022). Les narvals et les licornes se cachent-ils pour mourir? *Reseaux, 232233*(2), 169-196.

Ryan-Mosley, T. (2023, 4 October). How generative AI is boosting the spread of disinformation and propaganda. *MIT Technology Review*. https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/

Scott, J. 2021. Servicenow closes $230 Million USD Acquisition of Montreal's Element AI, *Betakit*, January 15, 2021. https://betakit.com/servicenow-closes-230-million-usd-acquisition-of-montreals-element-ai/

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society, 4*(2), 205395171773810. https://doi.org/10.1177/2053951717738104

Seaver, N. (2018). What Should an Anthropology of Algorithms Do? *Cultural Anthropology, 33*(3), 375–385. https://doi.org/10.14506/ca33.3.04

Senneville, M. (2021). Reconfiguration des liens de collaboration entre acteurs industriels et universitaires de la recherche en intelligence artificielle à Montréal et à Toronto [Mémoire de maîtrise]. Institut national de la recherche scientifique.

Silcoff, S. 2020. 2020. "Element AI Sold for $230-Million as Founders Saw Value Mostly Wiped Out, Document Reveals." *Globe and Mail*, December 21, 2020. https://www.theglobeandmail.com/business/article-element-ai-sold-for-230-million-as-founders-saw-value-wiped-out/

Simon, F. M., Altay, S., & Mercier, H. (2023, 18 October). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Mis/information Review*. https://misinforeview.hks.harvard.edu/article/misinformation-reloaded-fears-about-the-impact-of-generative-ai-on-misinformation-are-overblown/

Singhal, K. et al. (2023). Large language models encode clinical knowledge. *Nature, 62*, 172–180. https://doi.org/10.1038/s41586-023-06291-2

Statistics Canada. (2019). Newspaper Publishers, 2018. *Statistics Canada*. https://www150.statcan.gc.ca/n1/daily-quotidien/191129/dq191129d-eng.htm

Suchman, L. (2023). The uncontroversial 'thingness' of AI. *Big Data & Society, 10*(2), 20539517231206794. https://doi.org/10.1177/20539517231206794

Sutton, R. (2019). The bitter lesson. *Incomplete Ideas (blog), 13*(1).

Taylor, J., & Hern, A. (2023, mai 2). 'Godfather of AI' Geoffrey Hinton quits Google and warns over dangers of misinformation. *The Guardian*. https://www.theguardian.com/technology/2023/may/02/geoffrey-hinton-godfather-of-ai-quits-google-warns-dangers-of-machine-learning

Togelius, J., & Yannakakis, G. N. (2023). Choose your weapon: Survival strategies for depressed AI academics. *arXiv preprint arXiv:2304.06035*.

Vincent, J. (2019, 27 mars). Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI. *The Verge*. https://www.theverge.com/2019/3/27/18280665/ai-godfathers-turing-award-2018-yoshua-bengio-geoffrey-hinton-yann-lecun

Wang, J. (2018). Carceral Capitalism. *Semiotext(e)*.

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus, 109*(1), 121–136. https://www.jstor.org/stable/20024652

Zagni, G., & Canetta, T. (2023, 5 April). Generative AI marks the beginning of a new era for disinformation. *European Digital Media Observatory*. https://edmo.eu/2023/04/05/generative-ai-marks-the-beginning-of-a-new-era-for-disinformation/

Zuboff, S. (2020). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affair Books.