Introduction – Les réflexions philosophiques étudiantes sur l'intelligence artificielle

Keven Bisson, *Codirecteur de la revue Phares, Université Laval et* Romane Marcotte, *Codirectrice de la revue Phares, Université Laval*
Avec les commentaires de Jocelyn Maclure, *professeur titulaire à la Faculté de philosophie de l'Université Laval*

Au cours de la dernière décennie, les développements fulgurants de l'intelligence artificielle (IA) ont pris une place centrale dans le monde intellectuel. Le Québec est d'ailleurs un acteur privilégié de ces développements et des réflexions qui y sont liées : le Québécois Yoshua Bengio a gagné en 2018 l'un des plus importants prix dans le domaine pour son travail sur les réseaux neuronaux profonds (le prix Turing), des entreprises comme Google et Facebook viennent s'installer à Montréal, l'Observatoire international sur les impacts sociétaux de l'IA et du numérique s'est installé à Québec, etc. La Faculté de philosophie de l'Université Laval ne fait pas exception, et a augmenté considérablement son offre de cours touchant le sujet de l'intelligence artificielle. C'est dans ce contexte que s'inscrit ce numéro de la revue Phares, qui présente un aperçu du travail des étudiantes et étudiants en philosophie sur ce sujet, afin de stimuler les réflexions présentes et à venir sur ce thème. Parmi les multiples inquiétudes que soulève le développement effréné des algorithmes de l'IA, deux grandes problématiques intéressent les recherches philosophiques présentées dans ce dossier : la possibilité d'une conscience artificielle et les conséquences éthiques de l'utilisation et de la constitution de machines intelligentes.

D'abord, concernant la première problématique, plusieurs chercheur.es croient que la création d'une intelligence artificielle consciente est possible. Toutefois, le fait que l'être humain ait

des émotions et un corps, qui sont partie intégrante de son expérience consciente, semble être un obstacle important à l'atteinte de cet objectif. Est-il possible de créer une IA qui dépasserait cette difficulté, c'est-à-dire qui serait dotée d'un point de vue subjectif sur le monde qui soit comparable au nôtre ? C'est à cette question que s'attardera le premier texte de notre dossier. Cet article est avant tout un refus de réduire la conscience humaine à une somme de processus neuronaux, de penser notre esprit comme un simple logiciel implanté dans la matière du cerveau - comme le font certaines théories cognitivistes sur lesquelles s'appuient les tenants de l'IA forte. Lorsqu'on aborde la conscience sous un autre angle, c'est-à-dire en la considérant comme le lieu de nos vécus des phénomènes du monde, la possibilité qu'une machine puissent un jour posséder une conscience comparable à la nôtre se voit considérablement réduite. Pour sa démonstration, Geneviève Fréchette détaille d'abord les caractéristiques qui font de notre conscience une conscience phénoménale en s'appuyant sur la phénoménologie de Husserl. Les principales caractéristiques retenues seront entre autres : l'ancrage dans un corps matériel, la possession d'une volonté, la possession d'un point de vue subjectif, la capacité d'expérimenter des vécus intentionnels et celle d'opérer une synthèse des vécus. L'auteure en vient finalement à déterminer l'impossibilité pour une IA forte de posséder une conscience qui puisse être dite phénoménale, puisqu'elle serait incapable d'appréhender une expérience depuis un point de vue proprement subjectif, relevant du domaine du sens.

Si on considère cependant possible la création d'une IA consciente, un autre problème se pose : jusqu'à maintenant, l'être humain légitimait son statut moral particulier par la possession d'une conscience. Comment alors considérer moralement une entité artificielle qui possèderait une intelligence égale - voire supérieure - et une conscience semblable à celle de l'être humain ? Cette question est particulièrement importante puisqu'on envisage la conception de robots qui entreraient en relation avec des êtres humains, notamment pour accomplir à leur égard diverses tâches de soin. Dans cette dernière optique, on songe notamment à créer des sexbots, c'est-à-dire des robots conçus pour avoir des relations

intimes et sexuelles avec les êtres humains. La sexualité constitue déjà un aspect de la vie humaine plein de nuances et de paradoxes. La possibilité de la conception de tels robots ainsi que les implications éthiques de leur mise en marché nécessitent donc une réflexion approfondie, ce à quoi s'attèle le second texte de notre dossier. Dans cet article, Samuel Nepton confronte directement l'affirmation de Joanna Bryson selon laquelle les robots devraient être considérés comme des esclaves et traités comme de simples objets. Selon lui, que l'on songe à la construction de sexbot « faibles » (des robots qui ne feraient qu'imiter des sentiments), ou de sexbots « forts » (des robots véritablement sensibles et conscients d'eux-mêmes), ces machines constitueraient un véritable angle mort aux thèses de Bryson. Cet angle mort se décline en deux catégories. D'un côté, la réduction des sexbots au statut d'objet risquerait d'aggraver la situation de vulnérabilité des humains avec lesquels ils doivent entretenir une relation de soin. De l'autre côté, dans l'optique où ces robots seraient conçus comme des êtres sensibles, ils seraient susceptibles de souffrir et donc auraient besoin de protection.

L'auteur du texte nous fera d'abord remarquer l'importance que ces machines ne soient pas de simples jouets sophistiqués de masturbation, puisque leur conception naît d'un besoin que certaines personnes ont de nouer de véritables relations. C'est donc au niveau relationnel qu'il faudra juger le statut de ces robots - une approche possible grâce au travail de Mark Coeckelberg. Une telle approche permet de dégager trois manières dont la réduction du robot au stade d'objet rendrait les « propriétaires » des sexbots, et les sexbots eux-mêmes, vulnérables. D'abord, certains humains qui seraient incapables de trouver l'amour, souvent des personnes marginalisées, seraient encore plus isolés socialement du fait qu'on les considèrerait en relation avec de simples « objets ». Dans un tout autre ordre d'idée, certains sexbots personnalisés pourraient permettre à des êtres humains au comportement sexuel déviant (on nomme par exemple le cas de la pédophilie) d'exacerber leurs tendances pathologiques. Finalement, dans le cas où l'on concevrait des sexbots forts, ceux-ci seraient capables d'expérimenter des sentiments comme l'amour et

la tendresse, mais seraient aussi sujet à la souffrance. Pour contrer ces problèmes, l'auteur propose l'adoption d'un cadre légal minimal.

L'utilisation actuelle de l'IA concerne également des aspects plus près de notre réalité, comme l'augmentation de l'efficacité de nos institutions par le traitement de données massives orientant la prise de décision. L'IA traite en effet déjà des demandes de libérations conditionnelles, ou encore de l'octroi de prêts dans certaines banques. L'IA présente deux avantages par rapport à l'être humain pour accomplir ce type de tâches. D'un côté, elle est plus « intelligente » que l'être humain, puisqu'elle peut considérer plus de facteurs complexes en même temps pour prendre une décision. D'un autre côté, le caractère « artificiel » de cette intelligence s'oppose au caractère « naturel » de l'intelligence humaine, teintée d'émotions, d'un vécu particulier et de préjugés. Cependant, contrairement à ce qu'on pouvait espérer, la courte expérience que nous avons de cette utilisation de l'IA montre que les algorithmes sont également empreints de discriminations. Dans le troisième texte du dossier, Sandrine Charbonneau s'attarde à cette discrimination algorithmique. En plus de nous montrer par plusieurs exemples l'ampleur des conséquences que ces algorithmes biaisés peuvent avoir sur la vie des individus, l'auteure explore les diverses causes qui pourraient être à leur origine, notamment le manque de diversité et de sensibilisation chez les programmeurs et programmeuses. À ce problème s'ajoute celui de l'opacité des algorithmes de deep learning, dont le fonctionnement est considéré comme un secret d'affaires. En effet, les recherches qui souhaiteraient repérer les articulations problématiques de ces algorithmes sont difficiles à effectuer en raison de cette impossibilité à accéder aux algorithmes. En réponse à ces problématiques, l'article propose plusieurs pistes d'encadrement d'utilisation des algorithmes de deep learning, notamment la mise en place d'une procédure rigoureuse d'audit afin de tester ces programmes avant de les lancer sur le marché. Il souligne aussi l'importance de sensibiliser les programmeurs et programmeuses responsables des collectes de données à la source du fonctionnement des algorithmes, afin qu'ils et elles soient conscient.es de l'impact de leurs biais dans leur travail. Finalement,

l'article nous met en garde contre une confiance aveugle dans ces programmes d'intelligence artificielle, et insiste pour que la décision finale reste prise par un humain.

Les algorithmes ne se restreignent pas seulement à l'accomplissement de tâches que nous faisions, ils peuvent en accomplir de nouvelles que nous n'aurions jamais pensé possibles si elles devaient être faites par des êtres humains. En effet, les algorithmes permettent de sélectionner le contenu présenté sur différentes plateformes publiques afin de le rendre plus personnalisé pour son utilisateur ou son utilisatrice. Or, l'organisation du contenu de ces plateformes très fréquentées (pensons aux réseaux sociaux), n'est pas sans influencer notre manière de voir le monde, et notamment nos opinions politiques. Le dernier texte du dossier explore l'impact de cette influence dans le contexte de la polarisation des opinions politiques aux États-Unis. En effet, Éric Gagnon remarque que le fossé idéologique entre partisans démocrates et républicains n'a cessé de se creuser au cours des dernières années, et constate que la plateforme Facebook a été un des milieux à l'origine de cette division. Afin de développer le lien entre les algorithmes de ce réseau social et la progressive radicalisation des opinions politiques américaines, l'article explore le fonctionnement de la raison humaine. En suivant les thèses d'Hugo Mercier et Dan Sperber, Éric Gagnon se trouve en mesure de déterminer que les environnements les plus propices à l'exercice de la raison ainsi qu'au déroulement d'une délibération saine sont avant tout des lieux dits « hétérogènes ». Ces milieux se caractérisent avant tout par une pluralité d'opinions, qui ne permettent pas à l'individu de se conforter dans ses propres croyances, et qui le poussent à formuler des arguments rigoureux. Dans un contexte contraire, l'homogénéité d'un milieu rend la raison paresseuse, et, en raison du biais du parti pris, moins apte à produire des arguments solides. Sachant cela, l'article fait remarquer que les algorithmes utilisés sur les plateformes des réseaux sociaux sont précisément conçus pour nous exposer à du contenu politique adapté à nos préférences. Ces plateformes, telles qu'elles sont utilisées jusqu'à maintenant, ne constituent donc pas des lieux de délibération, mais de radicalisation, et leur fonctionnement

nuit au développement critique des citoyens. On souligne d'ailleurs dans le texte les liens statistiques entre l'utilisation de Facebook de certains élus et la radicalisation de leurs partisans. À partir de cette conclusion, le texte propose plusieurs solutions d'encadrement de l'utilisation des algorithmes sur les réseaux sociaux. Il prône notamment la mise en place de mesures gouvernementales qui exigeraient de ces programmes qu'ils nous présentent un contenu politique aussi diversifié que rigoureux. En attendant de telles réformes, l'auteur de l'article invite son lectorat à diversifier par lui-même le flux d'informations auquel il est confronté au quotidien.