

L'identité humaine au défi de l'intelligence artificielle

L'autonomisation des machines, d'une part, l'hybridation croissante des humains avec la machine, d'autre part, questionnent l'idée d'une spécificité humaine irréductible.

Des machines éthiques ?

La première zone de porosité entre humains et machines s'établit autour de la question de l'idée de machine éthique. En effet, une façon radicale d'aborder les questions soulevées par l'éventuelle délégation de décisions à des machines autonomes (intelligence artificielle) est d'envisager que de rendre les machines « éthiques » serait une solution aux problèmes évoqués plus haut dans ce rapport. Une telle piste de réflexion est liée à la question de savoir s'il est même possible de formaliser une éthique³⁴ afin de la programmer dans une machine. Autrement dit, **peut-on automatiser l'éthique ?** Ce problème est apparu au cours des débats comme l'un de ceux retenant particulièrement l'attention de la communauté des chercheurs en intelligence artificielle. Gilles Dowek (CERNA) l'a ainsi souligné lors de la journée d'étude organisée au Collège des Bernardins le 20 septembre 2017.

Le fameux dilemme du tramway est très souvent évoqué à l'occasion de réflexions portant sur ce problème. On sait que ce dilemme met en scène un tramway sans freins dévalant une pente ; le tramway arrive devant un embranchement ; selon qu'il s'engage sur l'une ou l'autre des deux voies, il tuera une personne ou bien plusieurs. Dès lors, quelle devrait être la conduite d'une personne ayant la possibilité de manœuvrer l'aiguillage et donc de choisir, pour ainsi dire, l'un des deux scénarios possibles ? L'intérêt de cette expérience de pensée est qu'elle peut donner lieu à toute une gamme de variations : qu'en est-il si la personne seule attachée à l'une des deux voies se trouve être un proche parent ? Si les personnes sur l'autre voie se trouvent être 5 ou bien 100 ?

On voit aisément comment ce dilemme peut être adapté à l'hypothèse de voitures autonomes qui seraient mises en circulation prochainement : selon quels principes une voiture placée dans une situation de dilemme éthique de ce type devrait-elle « choisir » de se comporter ? Le dilemme du tramway a l'intérêt de mettre en évidence le fait que

différents choix « éthiques » sont possibles. Dès lors que des dilemmes de ce type auraient été anticipés au stade du développement du système, il serait bien sûr possible de leur donner une réponse. Mais précisément, **la spécificité de l'éthique n'est-elle pas de concerner des situations inédites, impliquant éventuellement des conflits de valeurs dont la solution doit être élaborée par le sujet** (pensons à Antigone, prise entre éthique familiale et éthique civique) ? N'est-elle pas de s'élaborer toujours en situation ? Dès lors l'hypothèse d'une formalisation de l'éthique n'est-elle pas quelque peu illusoire ? À tout le moins, elle implique une conception implicite de l'homme qui n'a rien d'évident.

Retenons du moins que, pour l'heure, des expressions comme « éthique des algorithmes » ou « algorithmes éthiques » ne doivent pas être prises au pied de la lettre et comportent une part d'anthropomorphisme revenant à attribuer des capacités humaines à des machines. Certains considèrent qu'elles sont susceptibles de fausser un débat qui devrait se concentrer sur les exigences à l'égard des hommes qui conçoivent, entraînent, déploient et utilisent les systèmes algorithmiques et d'intelligence artificielle.

Elles ne constitueraient alors qu'une métaphore commode mais à ne pas entendre littéralement. À l'inverse, comme le rappelle par exemple Gilles Dowek, on peut considérer comme légitime le recours à ce type de métaphores dans la mesure où elles reviennent à prendre acte de l'autonomie croissante de ces systèmes et de la nécessité de formaliser, autant que faire se peut, une éthique et de la programmer dans des algorithmes. Quoi qu'il en soit, même si une éthique en tant que telle pouvait être encodée dans une machine (c'est-à-dire si cette dernière avait la possibilité de ne pas seulement répondre d'une certaine façon à une situation éthique envisagée à l'avance lors de son développement mais bien d'aborder des situations nouvelles en leur appliquant un raisonnement éthique), le choix du type d'éthique à encoder resterait bien, en dernière analyse, du ressort de l'homme. Le vrai enjeu est alors de s'assurer que les choix éthiques faits au stade du développement ne font pas l'objet d'une confiscation par « une petite caste de scribes » (Antoine Garapon). L'échelle de déploiement des algorithmes à l'heure du numérique en fait une question démocratique essentielle.

³⁴ C'est-à-dire une règle générale d'évaluation de la conduite à adopter face à toute situation – éthique déontique ou éthique conséquentialiste – ou un corpus de règles remplissant la même fonction – éthique kantienne, éthique bouddhiste, etc.

L'hybridation de l'homme et de la machine : repenser l'identité humaine ?

Une façon d'envisager la question éthique appliquée aux algorithmes et à l'intelligence artificielle peut être de confronter ces derniers à l'affirmation – présente à l'article premier de la loi Informatique et libertés – selon laquelle l'informatique « ne doit pas porter atteinte à l'identité humaine ».

Les pages précédentes ont abordé des problèmes liés à la façon dont l'homme agence son action avec des artefacts, question ancienne mais renouvelée par l'émergence d'artefacts dotés d'une « autonomie » croissante à l'heure des algorithmes et de l'intelligence artificielle³⁵. Ces propos soulignent en effet que le développement de ces technologies, selon la manière dont il s'opérera, peut affecter l'une des composantes de l'identité et de la dignité humaines, à savoir sa liberté et sa responsabilité. La montée en puissance d'une forme d'« autonomie » machinique doit bien sûr être fortement nuancée. Gérard Berry, professeur au Collège de France et titulaire de la chaire « Algorithmes, machines et langages » rappelle ainsi : « un jour, nous dit-on, les machines parleront et seront autonomes, le numérique donnera naissance à une nouvelle forme de vie. La date pour l'autonomie des machines et leur capacité de parole créative, personne ne la donne, et je ne la connais pas, loin de là. Surtout, de quelle vie parlons-nous ?³⁶ ». Néanmoins, on pourrait se demander si la trajectoire technologique d'ores et déjà à l'œuvre ne devra pas conduire à questionner la pertinence de la notion même d'« identité humaine », dans la mesure où celle-ci implique une séparation étanche entre humain et non-humain. La question du « droit des robots » d'ores et déjà soulevée par des juristes et récemment examinée par le Parlement européen (rapport Delvaux) a pour horizon ce brouillage possible des frontières de l'humain. À de tels arguments post-humanistes, la tradition humaniste pourrait certes rétorquer que l'autonomie machinique n'est

aujourd'hui qu'un leurre, une métaphore destinée à styliser un objet complexe et masquant finalement une responsabilité et une action humaines certes diluées, éclatées, mais bien réelles.

Si une première hybridation entre l'homme et la machine s'opère au plan de l'action, la réflexion devra aussi nécessairement s'élargir à l'avenir pour prendre en compte l'hybridation physique parfois annoncée entre algorithmes, humains, voire animaux (avec l'adjonction d'implants intelligents et communicants). Cette hybridation physique est une étape supplémentaire de l'évolution déjà à l'œuvre dans l'interaction permanente qui nous lie d'ores et déjà à une foule de processus algorithmiques.

Enfin, ce thème d'une subversion éventuelle de la frontière entre l'homme et les choses (ou plutôt, entre l'homme et la machine) trouve déjà une réalité extrêmement concrète au plan phénoménologique dans certaines tentatives récentes d'applications de la robotique qui s'illustrent d'abord dans la forme humaine donnée aux robots. On pense ici au robot Pepper de la firme Aldebaran, destiné à être déployé dans des espaces commerciaux pour interagir avec les clients. Surtout, et ceci concerne directement le sujet des algorithmes et de l'intelligence artificielle, **tout un champ de recherche vise à créer des robots empathiques capables de percevoir les émotions des humains** (par l'analyse du visage, de la voix, etc.) de façon à s'adapter à leur interlocuteur. La première question posée par ces recherches est évidemment celle de la limite entre, d'une part, les apports bénéfiques d'une intelligence artificielle capable de comprendre et de s'adapter aux états émotionnels de ses interlocuteurs et, d'autre part, une forme de manipulation appuyée sur une ingénierie technique capable d'exploiter les vulnérabilités affectives des personnes³⁷. La seconde question, connexe à la première, est celle de savoir dans quelle mesure la capacité d'illusion propre à ces technologies et l'asymétrie qui existera entre ces robots et les personnes dont ils analyseront les émotions les rendent moralement acceptables ? Sherry Turkle, professeure au MIT, souligne ainsi que les êtres humains ont une grande propension à attribuer aux robots une subjectivité et une sensibilité³⁸. Or, la tentation est forte pour des sociétés vieillissantes de confier de plus en plus le soin des personnes âgées à ce type de robots. En France, Serge Tisseron développe une réflexion critique sur ces technologies³⁹. Quelles que soient les réponses apportées à ces questions, il semble essentiel qu'elles n'occultent nullement la dimension politique et de choix de société que recèle le fait de recourir aux robots plutôt que d'investir dans d'autres types de ressources (temps, ressources en personnel, etc.) pour l'accompagnement des membres vulnérables de nos sociétés.

Le développement de ces technologies peut affecter l'une des composantes de l'identité et de la dignité humaines, à savoir sa liberté et sa responsabilité

³⁵ La question de l'hybridation entre l'homme et des artefacts n'est pas nouvelle : les algorithmes participent au modelage de notre identité de la même façon que – Socrate le remarquait déjà dans le *Phèdre* de Platon – l'écriture affecte notre capacité de mémorisation et constitue un artefact muet, incapable de la moindre explication. Que l'idée d'une « identité humaine » strictement distincte des objets soit remise en cause n'implique ainsi pas nécessairement une nouveauté radicale.

³⁶ Gérard Berry, « Non, l'intelligence artificielle ne menace pas l'humanité ! », interview donnée au Point, 18 mai 2015.

³⁷ Une problématique très similaire à celle soulevée par les logiciels de communication politique censés adapter le message du candidat aux attentes de chaque individu ciblé et profilé.

³⁸ Sherry Turkle, *Seuls ensemble*, Paris, L'Echappée, 2015 [2012].

³⁹ Serge Tisseron, *Le Jour où mon robot m'aimera. Vers l'empathie artificielle*, Paris, 2015.