

ÉTHIQUE ET IMPACT SOCIÉTAL

ÉTHIQUE ET IMPACT SOCIÉTAL

Comme nous l'avons vu dans les bulletins précédents, l'IA, déjà présente dans notre quotidien, permet de nous faciliter des actes de la vie quotidienne. Elle a également comme atout de permettre à l'humain de renforcer ses savoir-faire. Regardons de plus près les avancées en IA concernant les émotions humaines et celles qui pourraient être utilisées dans le système éducatif pour comprendre les enjeux éthiques et sociétaux.

L'IA ET LES ÉMOTIONS

La capacité de décodage des actions humaines par une machine permet à celle-ci d'induire des émotions et proposer un stimulus adéquat (une autre chanson chez Deezer, une activité chez Tega).

Tega : Il s'agit d'un petit robot à destination des enfants. Il a pour but de les accompagner dans leurs apprentissages et de s'adapter à leur comportement. Il est aussi capable de lire les émotions sur les visages et, en fonction de ce qu'il perçoit, agit pour stimuler l'enfant. Par exemple, lors d'un exercice, il peut observer l'ennui et répondre par des encouragements.

Deezer : L'application de streaming musical est parvenue à concevoir un algorithme capable de reconnaître l'énergie d'une chanson (triste, joyeuse, mélancolique...), l'objectif étant de connaître l'humeur de l'utilisateur pour lui proposer une playlist adaptée.

Comment ça marche ?

Pour que le robot puisse reconnaître les émotions, de très nombreuses images de visages ont été intégrées dans la machine pour qu'elle puisse les reconnaître et les associer à une étiquette (joie, tristesse, etc.). Elle pourra ensuite analyser les caractéristiques des expressions du visage des enfants pour les associer à une catégorie enregistrée dans le système et agir en conséquence. Deezer, quant à lui, peut analyser le rythme des chansons et les paroles. Il est donc capable de catégoriser la chanson. Dans l'exemple de Deezer, c'est le choix de la chanson qui permet au système de considérer que l'utilisateur est de telle ou telle humeur pour ensuite adapter la playlist.

Peut-on catégoriser les émotions humaines ?

Pour répondre à cette question, il est nécessaire de distinguer les émotions des sentiments. Les sentiments sont des expériences mentales privées qu'il est donc possible de dissimuler. Les émotions, en revanche, sont publiques et sont intégrées au mouvement se manifestant dans le corps humain. Le processus émotionnel se façonne à partir de différents facteurs comme le contexte, l'humeur ou encore l'expérience (Claudon et Weber, 2009). Aussi, le processus émotionnel est singulier, un même stimuli peut faire réagir une même personne différemment, sur des temps distincts, en fonction de différents facteurs (Damasio, 2010). Il est donc possible de catégoriser une émotion en observant les réactions du corps. Cependant, il n'est pas possible pour le moment de comprendre comment l'émotion se construit ni ce qu'elle signifie pour la personne.

Agir en s'appuyant uniquement sur ces catégorisations pourrait restreindre notre champ des possibles. Par exemple, un système sera-t-il capable de différencier un rire de joie, un rire de surprise, un rire de colère, un rire ironique ou tous ces rires seront-ils étiquetés « joie » ? Le risque ne serait-il pas de retirer l'aspect sensible, affectif de l'émotion qui façonne l'être humain ?

Il pourrait être réducteur de distinguer les émotions et de les cantonner à des catégories telles que la peur, la tristesse, la joie, la colère, qui sont certes les émotions fondamentales, mais qui n'interviennent ni seule, ni à un même degré lors de leurs apparitions (Devillers, 2017).

LES LIMITES DE L'IA

Les catégorisations prédéfinies par l'homme (IA supervisé) ou par la machine elle-même (IA non supervisé) présentent des risques de récurrences qui pourraient amener l'individu dans une boucle, c'est-à-dire qu'une émotion, une réaction ou un acte engendrerait une action récurrente.

Prenons comme exemple l'ennui qui peut être comblé par un robot éducatif capable de lire l'ennui sur le visage d'un enfant et qui pourrait répondre en lui proposant automatiquement une activité. Ce sentiment est-il négatif dans toutes ses dimensions ? Existe-t-il une réponse type à cette émotion ?

Roger Teboul [2005], par exemple, présente l'ennui, pour les enfants, comme la découverte de leur capacité à être seul. Celle-ci amène également l'enfant à développer son imaginaire et sa conscience de soi. À l'inverse, l'ennui peut aussi l'amener vers une perte d'intérêt envers son environnement. Parce qu'il est nécessaire que l'IA préserve la multiplicité de réponses face à une action ou une réaction humaine, dans quelle mesure l'IA pourrait-elle percevoir ces nuances ?

INÉGALITÉS, STÉRÉOTYPES ET PRÉJUGÉS

L'IA peut également provoquer des inégalités. Pour le comprendre, nous allons présenter les deux premières étapes de la conception d'un algorithme.

La première étape de la conception d'un algorithme est l'apparition d'une idée d'un prescripteur, par exemple, qui n'a pas de connaissance particulière en informatique. Bernheim et Vincent [2019] proposent l'exemple d'un directeur de master qui souhaiterait créer un algorithme pour trouver le meilleur étudiant des années inférieures. Ce prescripteur se tourne ensuite vers un développeur pour qu'il élabore le système (l'algorithme). À cette étape, les premiers biais peuvent apparaître. En effet, les deux scientifiques expliquent qu'à ce moment, c'est le développeur qui va faire le choix des critères de sélection de l'étudiant. Il pourrait se baser sur les résultats finaux de celui-ci, sur sa moyenne générale ou encore sur sa progression. Dans cet exemple, nous pouvons observer que la subjectivité/le point de vue du développeur donne une orientation à l'algorithme.

Le second exemple présenté concerne la seconde étape de la conception de l'algorithme. Il s'agit de l'étape d'apprentissage, c'est-à-dire le temps où le superviseur (l'informaticien, le data scientist, etc.) sera amené à entraîner la machine pour qu'elle puisse, dans un second temps, calculer et catégoriser en autonomie. Ainsi, il a été démontré que certains systèmes peuvent être discriminants

[Bernheim et Vincent, 2019]. Cet effet peut être dû aux types de profils intégrés dans les bases de données, lesquelles seront utilisées lors de la phase d'apprentissage. Dans certaines bases de données, il existe une surreprésentation d'un profil. Concrètement, cela peut amener un système de reconnaissance faciale à reconnaître une femme noire et pas une femme blanche ou encore un système de tri de CV qui considérerait que le bon profil est de l'un des deux sexes et donc mettre automatiquement de côté ce que le système considèrera comme un mauvais profil.

Suite à ce constat, de nouveaux systèmes ont été créés pour sonder les bases de données sur les disparités (âge, sexe, ethnologie...). Aussi, des bases de données standards à profils variés sans surreprésentations sont de plus en plus mises à disposition pour l'entraînement de certains systèmes. L'éthique et les impacts sociétaux sont importants à prendre en compte pour accompagner les avancées technologiques et ne pas répondre à une problématique tout en en créant une autre.

IMAGINONS L'AVENIR

Avec ses limites et ses risques, l'IA serait une avancée spectaculaire qui peut avoir de nombreux bénéfices dans notre société. Il nous semble nécessaire de relever une nouvelle fois l'importance de la collaboration entre l'IA et l'humain, et non la création d'une forme de compétition. Nos modes de vies, nos pratiques, et même l'organisation de la société évoluent et continueront d'évoluer au rythme des avancées technologiques. Il est essentiel, comme le montre le rapport Villani [2018] ou encore de nombreux chercheurs, de ne pas uniquement se focaliser sur les fonctionnalités des nouveaux outils. Il est important de se questionner sur les impacts sociétaux et humains dès la création d'un nouveau système d'IA pour favoriser l'égalité, le respect de la vie privée, l'intégrité et l'épanouissement humain. Il s'agit d'un défi majeur pour la société d'aujourd'hui et de demain.

RÉFÉRENCES

- Bernheim Aude, Vincent Flora, *L'Intelligence artificielle, pas sans elles !*, Paris, Belin, 2019.
- Claudon Philippe, Weber, Margot, « *L'émotion. Contribution à l'étude psychodynamique du développement de la pensée de l'enfant sans langage en interaction* », *Devenir*, vol. 21, n° 1, 2009, p. 61-99.
- Damasio Antonio, *L'Autre moi-même. Les nouvelles cartes du cerveau, de la conscience et des émotions*, Paris, Odile Jacob, 2012.
- Devilleurs Laurence, *Des robots et des hommes. Mythes, fantasmes et réalité*, Paris, Plon, 2017.
- Heudin Jean-Claude, *Intelligence Artificielle. Manuel de survie*, Science eBook, 2017.
- Teboul Roger, « Plaidoyer pour l'ennui », *La lettre de l'enfance et de l'adolescence*, vol. 60, n° 2, 2005, p. 25-30.
- Villani Cédric, *Donner un sens à l'intelligence artificielle. Pour une stratégie nationale et européenne*, Conseil national du numérique, 2018 [en ligne].