

## Ethique de l'Agent intelligent

Les questions concernant l'éthique sur l'IA peuvent se répartir sur deux axes :

- Est-ce que l'utilisation que l'on fait d'un système intelligent est éthique ?
- Est-ce que l'agent autonome conçu par l'Homme est moral/éthique ?

Qu'est ce qui fait la distinction entre ces deux problématiques ? Nous pensons que c'est le libre arbitre de l'agent qui fait la différence. En effet, un agent ne disposant d'aucun libre-arbitre prendra des décisions seulement en fonction de sa programmation, c'est plutôt le sens moral du programmeur qui entre en jeu, alors qu'un agent disposant d'un libre arbitre pourra réellement prendre une décision en considérant ses propres expériences et valeurs.

Nous allons donc nous intéresser à la seconde problématique.

Tout d'abord même si l'agent dispose d'un libre-arbitre, il faudra quand même lui donner un ensemble de règles éthiques de départ. Quelles seront ces règles ? S'apparenteront-elles à celles établies par Isaac Asimov ? Existera-t-il un consensus sur le sujet ou cela sera-t-il laissé à la discrétion de chaque programmeur ?

Ces agents devront-être capable d'évaluer et de comparer entre eux des principes éthiques. Comment fournir à ces agents les moyens de prendre la meilleure décision ?

Supposons ensuite qu'un agent intelligent dispose d'un libre-arbitre, même simulé, et d'un ensemble de règles de base, il est donc capable de remettre en question les principes qu'on lui a donné et pourrait donc se forger son propre système de valeur et évoluer hors de notre contrôle. On pourrait alors se poser la question, est-il raisonnable de laisser une telle possibilité à un système ?

S'il existe une communauté d'agents fonctionnant sur ce principe, finiront-ils par développer des principes communs, comme une « vraie » société, ou au contraire auront-nous des agents très hétéroclites sur le plan moral ? Seront-ils nécessairement nocifs l'humanité ?