

HOMO SAPIENS INFORMATICS chronique de Gilles Dowek

Le dilemme moral de la voiture autonome

Programmer un robot qui sera confronté à des décisions complexes nous oblige à expliciter nos choix éthiques.



Un dilemme moral est une situation dans laquelle nous devons faire un choix qui nous mène, dans tous les cas, à commettre une mauvaise action. Par exemple, si nous devons faire le choix entre voler un médicament pour soigner un malade et ne pas aider ce dernier, nous avons deux mauvaises actions possibles : voler ou laisser souffrir le malade. Dans une telle situation, le problème qui se pose à nous n'est plus celui de faire le bon choix, mais de faire le moins mauvais.

Les réflexions sur le moins mauvais choix sont cependant parfois perçues comme un peu artificielles, car elles mènent à des conclusions difficilement utilisables en pratique.

Par exemple, imaginons qu'un groupe d'enfants traverse, sans prévenir, une route de montagne sur laquelle roule une voiture. Le conducteur n'ayant plus le temps de freiner, son choix est soit d'écraser les piétons, soit de jeter sa voiture dans le précipice. Face à un tel dilemme, s'il y a davantage de piétons sur la route que de passagers dans la voiture, le moins mauvais choix est sans doute de jeter sa voiture dans le précipice, afin d'épargner le plus grand nombre de vies possible.

Or peu de conducteurs réagissent effectivement ainsi. Et il n'est pas difficile d'expliquer cette incohérence entre leurs valeurs — qui devraient les mener à limiter le nombre de victimes de l'accident — et leurs actions : faiblesse de la volonté, prépondérance de l'instinct de survie, difficulté d'appréhender la situation dans sa globalité, brièveté du temps imparti à la décision, etc.

Toutefois, ces réflexions sont probablement moins artificielles aujourd'hui que naguère. De nombreux véhicules — ascenseurs, métros, drones, etc. — n'ont en effet plus de conducteur et d'autres — trains, avions, voitures — n'en auront plus dans un avenir plus ou moins proche. De ce fait, quand nous concevons un tel véhicule autonome (voir l'article pages 46 à 51), en particulier quand nous concevons son logiciel de pilotage, nous



SELON QUELS CRITÈRES un véhicule autonome devra-t-il prendre une décision juste avant un accident inévitable ?

devons décider de ce que fera ce véhicule quand il identifiera un groupe de piétons sur la route, s'il n'a plus le temps de freiner et s'il n'a d'autres possibilités que d'écraser les personnes ou de se jeter dans le précipice.

Ce choix doit désormais être explicite dans le logiciel de pilotage : ici comme ailleurs, programmer un ordinateur exige d'explicitier l'implicite.

Contrairement à la situation précédente, la décision ne doit pas être prise sur le

coup, mais au moment de la conception du logiciel de pilotage du véhicule, c'est-à-dire longtemps à l'avance. Nous avons donc tout le temps d'y réfléchir sereinement et il n'est plus possible d'invoquer la difficulté de décider dans le feu de l'action pour justifier une incohérence entre nos valeurs et nos actions.

En décidant de sauver les passagers de la voiture plutôt que les piétons, nous ne faisons, dans le premier cas, que commettre une mauvaise action. Dans le second, nous faisons plus : nous déclarons aussi publiquement notre intention de la commettre. Ce qui n'était qu'une incapacité de décider dans l'instant, une faiblesse de la volonté, etc., sans doute pardonnables, devient le plus froid des cynismes.

Cette obligation d'explicitier nos intentions dans les logiciels change donc complètement le rapport de la pensée éthique à l'action. Et il est possible qu'après des siècles d'incohérence entre nos belles valeurs et nos moins belles actions, l'évolution des techniques nous contraigne à les accorder. Cette explicitation nous mènera à changer nos actions, mais sans doute aussi nos valeurs. Maintenant que ces dernières nous engagent réellement, nous devons sans doute les définir avec plus d'honnêteté.

Et une telle mise en cohérence de nos valeurs et de nos actions porte un nom : cela s'appelle un progrès moral. ■

Gilles DOWEK est chercheur à l'Inria et membre du conseil scientifique de la Société informatique de France.