



Pour David Rudrauf, 44 ans, professeur à la Faculté de psychologie de l'Université de Genève: «Les machines véhiculeront quelque chose de l'humanité au-delà de l'humanité.» Mark Henley

«J'aimerais contribuer à créer de nouvelles espèces d'êtres artificiels»

ELISABETH ECKERT

elisabeth.eckert@lematindimanche.ch

ROBOTIQUE David Rudrauf, professeur à l'Université de Genève, a développé un modèle mathématique du comportement humain, qu'il transpose déjà sur des robots qui nous ressembleront.

Franchement, pourquoi avez-vous voulu nous modéliser?

Je me suis toujours passionné pour la compréhension scientifique de ce qu'est la conscience ou l'âme de l'être humain et qui se retrouve, sans doute, dans une certaine mesure, chez de nombreux animaux. Ce

modèle de la conscience incarnée que nous avons développé, après dix ans de recherches, a vocation à avoir des applications concrètes, par exemple pour comprendre et prédire des comportements pathologiques, dont l'autisme, et améliorer les approches diagnostiques et thérapeutiques sur la base de l'intelligence artificielle.

Mais encore...

Mon rêve d'enfant est réellement de créer des machines qui auront, à défaut d'une vraie conscience, des comportements qui ressembleront aux nôtres, dotés d'une intelligence artificielle qui repose sur les mêmes principes que les nôtres, avec une mémoire propre. Des robots qui apprendraient de nous et desquels nous apprendrions et avec lesquels nous partagerons



des liens affectifs et imaginaires, car l'imagination est au centre de ce modèle. Dès lors, puisque nous sommes là dans le domaine du réalisable, j'aimerais contribuer à créer de nouvelles espèces d'êtres artificiels, libres et autonomes, agissant sur la base de ce modèle de conscience et qui pourront un jour véhiculer quelque chose de l'humanité au-delà de l'humanité. C'est ma modeste métaphysique à moi.

Ce modèle mathématique du comportement humain que vous avez élaboré a-t-il déjà été implanté dans des robots?

En partie, oui. Dans le laboratoire de modélisation multimodale de l'émotion et du ressenti que je dirige et qui fait partie du Centre interfacultaire en sciences affectives de l'Université de Genève, nous travaillons, en collaboration avec le Centre universitaire d'informatique, avec des petits robots dotés d'une version encore très simplifiée de notre algorithme. Or, même si tout cela est encore à un stade très primitif, on observe déjà qu'ils manifestent des comportements collectifs d'attention conjointe, car ils sont capables de prendre la perspective des uns, des autres, de se mettre à leur place. Ainsi, si l'un des petits robots détourne soudain son attention du groupe sur un autre point d'intérêt, certains de ses congénères le suivent, regardent ce qui a suscité son intérêt. Notre algorithme a permis cela: ils interagissent, s'intéressent à la perspective d'autrui, évaluent leur environnement et prennent une décision d'action.

Votre modèle évoque également la possibilité de prédire les comportements humains. Sera-ce l'ère du film «Minority Report», où un ordinateur prévoit les actions futures d'un individu?

Sur ce point, ma réponse est catégorique: non, du moins pas exactement! Il faudrait qu'une machine ou qu'un outil soit capable d'avoir accès à l'ensemble de la mémoire d'une personne et de celles avec qui elle interagit, puis de simuler tout cela à un niveau de détails, de paramètres et de conditions initiales tel que cela sera, à mon sens, impossible, même dans un avenir lointain. Les images de science-fiction d'une multitude d'électrodes qui

scanneraient tout notre cerveau, vous pouvez oublier! Même si nous parvenions, vous ou moi, à nous modéliser au travers de grands traits de personnalité ou de caractéristiques autobiographiques, nous serions encore très loin de définir avec précision ce qui fait que vous êtes vous, Elisabeth, et moi, David, et de prédire nos moindres pensées et actions. Par contre, créer à partir de zéro des machines, des robots, qui vivraient sur un support autre qu'organique comme le nôtre, avec, pour but, d'en faire des entités propres et uniques, c'est tout à fait envisageable.

Vous êtes donc parvenu à nous «réduire» à une équation «E = mc²»?

Non, notre modèle ne repose pas sur une seule équation mais sur plusieurs, et est assez complexe. Deux grands principes le gouvernent, en réalité. L'être humain est à la fois une espèce prévisible, mais extrêmement complexe. Nous agissons en fonction de milliers de paramètres émotionnels et motivationnels qui font appel à la fois au désir, aux souvenirs, à la souffrance ou au besoin de sécurité, et vont justifier nos prises de décision et planifier nos actions. Nous sommes en permanence soumis à des choix, qu'ils soient importants ou pas. Dès lors, notre question était de savoir: comment décide-t-on, alors que d'innombrables paramètres, conscients ou inconscients, se télescopent sans cesse? Par exemple: «Je dois apporter un gâteau à mes enfants? Dois-je l'acheter ou le faire moi-même?» Nous nous mettons en permanence non seulement dans notre propre perspective, mais nous sommes capables d'imaginer le point de vue d'autrui qui nous conseillera ou déconseillera de faire telle ou telle chose. Pour prendre une décision, on passe son temps ainsi à se projeter, à imaginer, à recourir à ses souvenirs, à anticiper les conséquences. À nous balader entre le passé, le futur et le présent.

Ce tourbillon de la pensée peut nous rendre fou...

En effet. Dès lors et pour agir dans cette complexité, nous recourons à un «champ de conscience» gouverné par ce



qui s'appelle une «géométrie projective», qui s'inscrit non seulement dans le temps, mais également dans l'espace, et nous permet déjà de faire des choix. Notre esprit visualise et imagine en perspective un monde qui n'est pas en perspective. Quand on regarde des rails de chemin de fer, ils semblent se rejoindre plus ils se rapprochent de l'horizon. Cette construction projective en trois dimensions, que nous avons modélisée sur des bases mathématiques, joue un rôle fondamental dans la conscience et le comportement, et permet de faire le lien entre perception, imagination et action.

Et le second principe?

À la géométrie projective, nous avons associé le calcul de l'énergie libre, à travers laquelle un système autonome peut, selon ses préférences et ses croyances, mesurer et maximiser les probabilités d'obtenir ce qu'il veut de la manière la plus sûre, la plus optimale possible, comme l'a démontré mon collègue de l'Univer-

sity College London, Karl Friston. Le champ de conscience combine l'énergie libre et la géométrie projective, pour explorer le monde de façon active et satisfaire nos préférences, tout en imaginant les conséquences anticipées de nos actions. Ce calcul permet de simuler, grâce à la réalité virtuelle, et de prédire dans les grands traits les états de conscience et les comportements d'un individu. Nous avons ainsi découvert que si un agent artificiel est privé d'imagination projective, il se comporte comme une personne atteinte d'autisme, ce qui nous permet d'orienter la recherche sur l'importance de l'imagination et de ses mécanismes spécifiques dans la prise en charge de cette maladie.

*Le Pr David Rudrauf sera l'un des conférenciers du 2e Artificial Intelligence Geneva Summit (AIGS), présidé par Me Nicolas Capt. Sur le thème «Des machines et des hommes», il se tiendra le 5 octobre à la Maison de la Paix, à Genève.
Renseignement sur: www.aigs.ch.*

En dates

1973

David Rudrauf est né le 16 décembre à Boulogne-Billancourt.

1996

Bachelor en psychologie, puis master en neurosciences à l'Université Pierre-et-Marie-Curie (Paris VI).

2001

Doctorat en neurosciences et en sciences cognitives, puis assistant à l'University of Iowa.

2016

Professeur à la Faculté de psychologie et des sciences de l'éducation à l'Université de Genève.