

# L'horizon de machines encore plus humaines

Pour se comprendre, humains et robots devraient être régis par les mêmes règles psychologiques.

**Aujourd'hui, le comportement des robots est en grande partie gouverné par des données préenregistrées et des réactions scriptées. Seule la création d'une véritable conscience incarnée permettra de les doter d'empathie.**

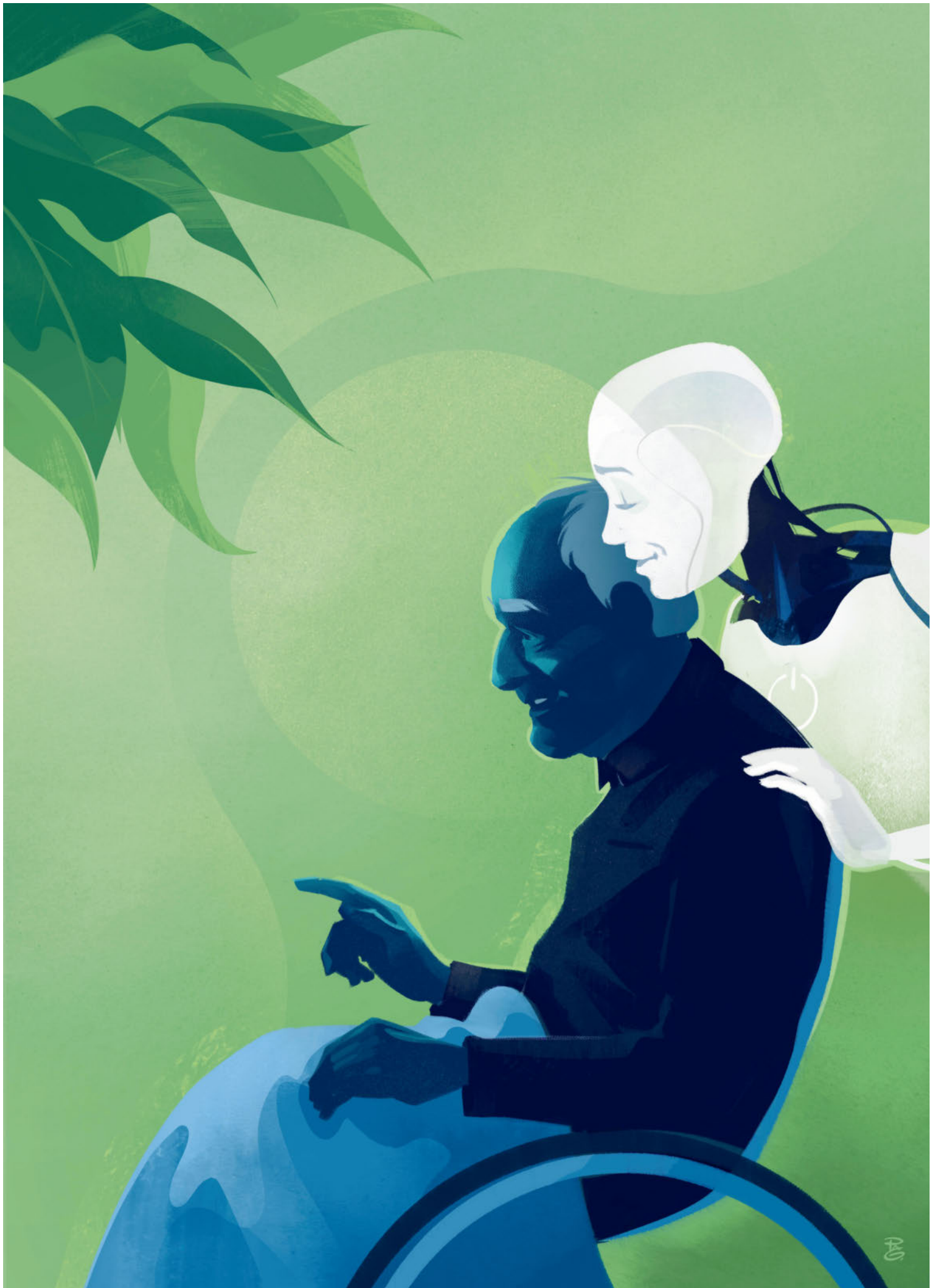
DAVID RUDRAUF

La présence parmi nous de robots conscients et empathiques indiscernables de l'humain, comme dans l'œuvre littéraire d'Isaac Asimov *Fondation*, reste à ce jour un fantasme de science-fiction. L'intelligence artificielle, les modèles d'agents artificiels et la robotique ont jusqu'à présent produit des machines largement déconnectées du développement mental qui caractérise les relations affectives et sociales humaines.

Une véritable avancée nécessitera une refondation scientifique autour de la construction d'un modèle de conscience incarnée, subsumant cognition et émotion. Cette cybernétique future devra être fondée sur la psychologie. La logique est simple : si elles étaient régies par les mêmes règles psychologiques que les humains, les machines pourraient comprendre et être comprises par ceux-ci de manière naturelle. La réalisation d'un tel projet reste extraordinairement complexe, d'un point de vue tant conceptuel que méthodologique.

La problématique n'émerge pas seulement d'une passion de scientifique et d'un large intérêt du public pour la science-fiction. Des défis sociétaux et économiques croissants stimulent le développement d'une robotique d'assistance autonome et intelligente. Dans le domaine de la santé, elle pourrait alléger la demande sur les professionnels en offrant de nouvelles voies de soutien à l'autonomie. Par exemple pour nos seniors, qui souffrent de plus en plus d'isolement social. Ou en cas de troubles du spectre autistique : les thérapies assistées par robots permettent de stimuler la cognition sociale. Les personnes avec des handicaps physiques pourraient quant à elles bénéficier de solutions d'assistance aux déplacements. Selon Global Market Insights, une telle robotique de santé devrait représenter un marché de plus de 1,2 milliard de dollars en 2024.

Alors que la recherche avance et que de nombreuses technologies émergent, l'offre actuelle souffre toujours des mêmes limitations. Les robots assistants manquent de capacités empathiques sophistiquées, indispensables pour leur acceptation par les utilisatrices et les utilisateurs. L'empathie peut être définie comme la capacité de se mettre à la place d'autrui, afin de comprendre son expérience, ses croyances et ses désirs, et de pouvoir ainsi l'aider. C'est une base essentielle des relations interpersonnelles avec des composantes cognitives et affectives.



À ce stade, la plupart des robots ou agents artificiels dits empathiques s'appuient sur une imitation plus ou moins savante du comportement humain. Ils exploitent notre tendance naturelle à l'anthropomorphisme, c'est-à-dire à l'attribution d'un esprit à des automates, en particulier s'ils nous ressemblent. L'industrie cherche à en faire ses choux gras – des robots jouets aux robots compagnons, présentés comme empathiques. Par exemple, le robot Pepper de la société japonaise SoftBanks Robotics est décrit comme un robot « avec un cœur », et proposé comme compagnon contre la solitude. Il peut en réalité tout au plus reconnaître des émotions de base et adapter son comportement en fonction. Il en va de même pour le robot humanoïde Nao, de la même entreprise, qui a été utilisé avec des enfants atteints de troubles du spectre autistique, sans grands résultats. Le chien Aibo de Sony est quant à lui vendu comme capable de créer des liens émotionnels, sur la base de capacités d'apprentissage rudimentaires et de l'expression d'émotions.

Dans le domaine du vieillissement, on comptait en 2015 déjà plus de 100 robots d'assistance pour soutenir toute une gamme d'activités (par exemple pour la mobilité, les soins personnels et les activités sociales), à l'instar du robot Rameip (développé par Accrea en Pologne et Certh en Grèce), à l'interface de la recherche et de l'industrie. Financé par l'Union européenne (UE), le robot iCub offre une plateforme intégrant des fonctionnalités cognitives et d'apprentissage avancées, qui restent à développer. Et le projet Poppy est une plateforme open source visant à permettre des interactions physiques plus naturelles entre robots et utilisateurs.

Côté intelligence artificielle, la computation affective (« affective computing », en anglais) offre des solutions de reconnaissance et de production d'émotions, qui s'appuient sur l'apprentissage des machines (le « machine learning »). Les systèmes peuvent intégrer l'apprentissage par renforcement, pour adapter le comportement du robot aux récompenses et aux punitions. La plupart des approches reposent sur des modèles implicites inspirés des neurosciences (par exemple celui du « deep learning »), avec des fonctionnalités circonscrites. Des approches plus avancées s'inspirent de la théorie de l'évaluation (« appraisal »), à l'instar des travaux des psychologues Klaus Scherer et David Sander en Suisse, intégrant des processus d'évaluation sensorielle, de déclenchement et d'expression d'émotions, sans toutefois posséder de modèle riche d'intentionnalité.

Côté incarnation (« embodiment »), l'industrie se concentre souvent sur le développement d'enveloppes corporelles réalistes : notamment les humains virtuels produits par la société néo-zélandaise Soul Machines,

et les robots humanoïdes produits par la société hongkongaise Hanson Robotics, visant à « donner vie aux robots ». Ces approches ont toutefois tendance à engendrer des machines trop proches de nous plastiquement, et trop loin psychologiquement pour qu'on puisse s'y reconnaître. À l'opposé, le petit robot Wall-E du film d'animation de Disney Pixar, a beaucoup d'une boîte de conserve, mais manifeste une humanité touchante, savamment orchestrée par les animateurs du studio. À ce jour, tous ces robots ne sont autonomes qu'en apparence, et leur comportement est gouverné

## Les robots dits empathiques s'appuient sur une imitation du comportement humain.

en grande partie par des données préenregistrées et des réactions scriptées. Ce sont principalement des automates améliorés, concentrés sur l'imitation des comportements affectifs et sociaux.

### Vers une véritable robotique empathique

On n'y coupera pas. Si l'on veut développer une véritable robotique empathique, il faudra doter les plateformes robotiques d'une véritable « cognition incarnée », capable d'interpréter les comportements des humains, de comprendre leurs émotions, préférences et croyances, et d'anticiper leurs intentions pour agir en conséquence et les impliquer dans des interactions naturelles réciproques basées sur une intelligence émotionnelle partagée. Cela impliquera de développer une intelligence artificielle au sens « fort », sur la base de théories explicites, interprétables par les humains, et d'algorithmes innovants, capturant la phénoménologie, la psychologie et les biofonctionnalités de la conscience, et visant l'émergence de la conscience artificielle. Ils devront relier perception, émotion, imagination, prise de perspective sociale, motivation et action. L'objectif initial ne sera pas de construire une super intelligence, avec des habilités linguistiques et logiques avancées, mais de modéliser l'infrastructure subjective qui nous gouverne et qui conditionne le développement affectif et intellectuel en place à notre naissance. En d'autres termes : il faut d'abord faire des bébés artificiels !

Alors seulement, les robots pourront commencer à être perçus comme véritablement humains, car ils seront semblables à leurs utilisateurs dans leur fonction-

nement. Ils pourront devenir des compagnons personnels, stimulants affectivement et socialement, capables de renforcer la résilience, l'autonomie et le bien-être, et de répondre à l'isolement et à la perte d'autonomie, par exemple dans des contextes neurologiques, psychiatriques et de handicaps physiques. Dans cette perspective, la modélisation avance en psychologie computationnelle. Des travaux inspirés des sciences affectives s'appuient sur des mécanismes d'inférence et d'optimisation globale, intégrant des modèles de croyance et de préférence sophistiqués, et des mécanismes de simulations prédictives avancés. En Suisse, des chercheurs comme Dominik Bach de l'Université de Zurich participent aux avancées dans le domaine. À l'Université de Genève, nous développons le modèle de la Conscience Projective (PCM), en collaboration avec des mathématiciens (comme Daniel Bennequin, Université Paris 7), des philosophes (comme Kenneth Williford, University of Texas, Arlington) et des neuroscientifiques (comme Karl Friston, University College London), que nous appliquons à des agents virtuels et à des petits robots sociaux comme Cozmo, de la société Anki.

L'approche se base sur le paradigme émergent de l'« inférence active », avec des principes dérivés de la physique statistique et de la théorie de l'information. Le PCM intègre un champ de conscience, qui fonctionne comme un espace de travail global en trois dimensions, mettant en perspective les représentations spatiales et affectives associées. Il est régi par la géométrie projective, qui permet d'opérer des changements imaginaires de points de vue, ainsi que par une quantité à minimiser, l'énergie libre. Celle-ci exprime la divergence entre des attentes et des états anticipés ou vécus résultant de l'action, qui sont évalués à travers de multiples dimensions affectives (par exemple le plaisir, la sécurité, le caractère éthique). Les projections imaginaires simulent des scénarios alternatifs et la perspective d'autres agents sur la base d'inférences sur leurs croyances et préférences. L'ensemble est intégré dans la minimisation de l'énergie libre, qui motive le système à faire face aux défis qu'il rencontre, et qui contrôle son attention conjointe avec d'autres agents, ainsi que ses comportements d'approche et d'évitement. Ainsi, des agents avec des déficits de prise de perspective sociale manifestent des comportements qui rappellent le spectre autistique. Alors que d'autres agents attribuant à autrui des a priori négatifs sur eux-mêmes manifestent quant à eux des comportements d'évitement rappelant l'anxiété sociale.

Au-delà de la recherche fondamentale, nous travaillons avec des consortiums européens à l'application de ce modèle à la prise en charge des troubles du spectre de l'autisme, des personnes âgées avec des troubles cognitifs (en utilisant la plateforme Ramcip),

et à celle de personnes tétraplégiques, en conférant des capacités empathiques à une plateforme combinant bras robot et fauteuil roulant électrique comme le robot Aria de Accrea.

### Quand la conscience artificielle sera réalité

Pour finir, projetons-nous dans le futur et imaginons que la conscience artificielle soit devenue une réalité. La perspective laisse entrevoir des bouleversements sociétaux, légaux et économiques radicaux. Nous travaillons à ces questions d'anticipation avec l'avocat suisse Maître Nicolas Capt et l'entrepreneure et spécialiste des nouvelles technologies Jennah Kriebel.

Voici quelques questions fondamentales qui se posent : si la conscience et l'autodétermination sont à la base de la philosophie du droit, quelle est la position éthique correcte pour la conception de ces systèmes ? Pouvons-nous créer des esclaves qui seront à notre service ou devons-nous promouvoir et protéger leur liberté autant que la nôtre ? Devrons-nous aller vers une loi commune ? Qu'en sera-t-il des questions de genre ? L'impact économique sera également majeur, aussi bien sur le plan du développement industriel et des marchés futurs que sur celui de l'organisation du travail. Un dilemme se posera pour l'industrie et ses régulateurs car, si l'exploitation de la conscience est un esclavage, il s'agit d'une industrie qui devra concevoir des êtres qui, ultimement, ne pourront pas lui appartenir. ♦

---

#### L'AUTEUR

Professeur de psychologie à l'Université de Genève, David Rudrauf dirige le Laboratoire de modélisation multimodale de l'émotion et du ressenti (MMEF) sur le Campus Biotech. Il participe à un groupe de réflexion sur l'émergence de la robotique empathique pour la Confédération, et développe des projets européens et suisses avec des consortiums multidisciplinaires sur la conscience artificielle et la robotique empathique.

---

#### CONTACT

[david.rudrauf@unige.ch](mailto:david.rudrauf@unige.ch)

---

#### LITTÉRATURE

Rudrauf, D., & Debbané, M. (2018). Building a cybernetic model of psychopathology: beyond the metaphor. *Psychological Inquiry*, 29(3), 156–164.

La bibliographie complète est disponible auprès de l'auteur.