

Les robots font de la science

Ross King

Des machines peuvent-elles émettre des hypothèses, réaliser des expériences et évaluer les résultats, sans l'intervention de l'homme ? Des prototypes montrent que oui.

Peut-on automatiser la recherche scientifique ? Il ne s'agit pas d'automatiser les expériences – cela se fait déjà –, mais de construire une machine qui découvrirait de nouvelles connaissances scientifiques. Depuis dix ans, mon équipe et moi développons un tel robot scientifique.

Pourquoi construire un robot scientifique ? Nous avons deux motivations. La première est de mieux comprendre la science. Le physicien Richard Feynman remarquait que « ce que je ne sais pas créer, je ne le comprends pas ». Dans la même ligne philosophique, tenter de construire un robot scientifique nous oblige à prendre des décisions concrètes reliant des objets abstraits et physiques, des phénomènes observés et théoriques, ainsi que nos différentes façons de poser des hypothèses.

Notre seconde motivation est technologique. Les robots scientifiques pourraient rendre la recherche plus productive et rentable. Certains problèmes scientifiques sont si complexes qu'ils nécessitent une importante quantité de travail : il n'y a pas assez de chercheurs pour tout faire. L'automatisation pourrait résoudre ce problème.

En science, l'informatique a connu d'importants progrès, par exemple l'automatisation des laboratoires à « haut débit » pour le séquençage de l'ADN et la recherche de nouveaux médicaments (avec les techni-

ques dites de criblage). Les ordinateurs effectuent même des opérations moins évidentes : ils analysent aussi des données et commencent à produire des hypothèses scientifiques. En chimie par exemple, les programmes d'apprentissage automatique aident à la conception de nouvelles molécules. L'intérêt d'un robot scientifique serait de combiner ces techniques pour automatiser l'ensemble de la démarche scientifique : formuler des hypothèses, concevoir et réaliser des expériences pour tester ces hypothèses, interpréter les résultats et recommencer le cycle jusqu'à ce que de nouvelles connaissances soient trouvées.

Sommes-nous capables de mettre au point un robot scientifique qui puisse vraiment accomplir cette démarche ? Les caractéristiques de deux robots que nous avons conçus, et de quelques autres dans le monde, suggèrent que oui.

Un robot scientifique nommé Adam

C'est dans les années 1960 et 1970 que des chercheurs de l'Université Stanford, aux États-Unis, ont pour la première fois tenté d'appliquer l'intelligence artificielle à la découverte scientifique. Ils ont conçu un programme informatique nommé DENDRAL pour analyser les données de spectrométrie de masse, et le programme

apparenté, Meta-DENDRAL, a été l'un des premiers systèmes d'apprentissage automatique. Les scientifiques voulaient créer des instruments capables de chercher des signes de vie sur Mars dans le cadre de la mission *Viking* de la NASA en 1975. Mais la tâche était trop importante pour la technologie de l'époque.

Depuis, des programmes tels que *Prospector* (pour la géologie), *Bacon* (pour l'exploration de terrains) et des successeurs plus récents ont automatisé certaines tâches proposant des hypothèses et des expériences à mettre en œuvre. Toutefois, la plupart de ces automates ne réalisent pas eux-mêmes les expériences. C'est pourtant une composante nécessaire si l'on veut que les systèmes à intelligence artificielle fonctionnent même de façon semi-indépendante.

Notre robot, Adam, n'est pas un humanoïde ; c'est un laboratoire automatisé qui remplit un petit bureau (voir l'encadré page 71). L'équipement comprend notamment un congélateur, trois robots pour manipuler les liquides, trois incubateurs, une centrifugeuse, et chaque composant est automatique. Adam a aussi un « cerveau » réalisant des calculs : c'est un ordinateur qui fait des raisonnements, contrôle les ordinateurs individuels actionnant le matériel, et compare les résultats des expériences à ceux prédits selon chaque hypothèse.

Adam réalise des expériences sur les mécanismes de croissance des microbes ; il sélectionne des souches microbiennes et des milieux de culture (des liquides nécessaires à leur développement), puis observe comment les souches se multiplient dans le milieu sur plusieurs jours. Le robot peut tester seul environ 1 000 combinaisons de souches et de milieux par jour. Nous avons conçu Adam afin qu'il étudie un domaine important de la biologie : la génomique fonctionnelle, à savoir les relations entre les gènes et leurs fonctions.

La première étude concernait la levure *Saccharomyces cerevisiae*, un micro-organisme constitué d'une seule cellule et utilisé pour faire du pain ou de la bière par exemple. Les biologistes s'intéressent surtout à cette levure en tant que « modèle » pour comprendre comment fonctionnent les cellules humaines. Les levures ont beaucoup moins de gènes que les cellules humaines et se multiplient rapidement et facilement. Bien que le dernier ancêtre commun des hommes et des levures ait vécu il y a environ un milliard d'années, l'évolution conserve de nombreux caractères, de sorte que l'on peut considérer que ce qui est vrai pour une levure l'est aussi pour une cellule humaine.

Adam a étudié les enzymes, des protéines qui catalysent, c'est-à-dire favorisent et accélèrent, des réactions biochimiques particulières : comment la levure les utilise-t-elle pour se multiplier et produire des déchets à partir de son milieu de culture ? Bien que les scientifiques étudient ces mécanismes depuis plus de 150 ans, ils n'en comprennent pas tous les détails. Ils connaissent de nombreuses enzymes produites par la levure, mais dans certains cas, ils ignorent les gènes qui les codent. Adam a pour objectif de découvrir ces gènes.

Pour faire des découvertes, Adam doit disposer d'importantes connaissances scientifiques. Nous avons programmé Adam avec un bagage approfondi de données sur le métabolisme et la génomique fonctionnelle des levures. Certains disent qu'Adam détient des informations plutôt qu'il n'a des « connaissances ». Le débat philosophique est ouvert. Mais nous pensons que le terme de connaissances est justifié, parce qu'Adam les utilise pour raisonner et guider ses interactions avec l'environnement.

La logique permet à Adam de représenter ses connaissances. Elle a été inventée il y a 2 400 ans pour décrire la connaissance

avec plus de précisions que ne le permet le langage. La logique est la façon la plus juste de représenter les données scientifiques et d'en échanger entre robots et êtres humains. De façon pratique, en intelligence artificielle, la logique peut aussi être un langage de programmation, ce qui permet aux connaissances d'Adam d'être interprétées comme un logiciel informatique.

Lui fournir des connaissances

Pour lancer les recherches d'Adam, nous l'avons programmé avec de nombreuses données. Par exemple, chez *S. cerevisiae*, le gène *ARO3* code une enzyme nommée 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase. Nous avons également fourni à Adam des données liées, par exemple que cette enzyme catalyse une réaction chimique transformant les composés phosphoénolpyruvate et D-érythrose 4-phosphate en 2-dehydro-3-deoxy-D-arabino-heptonate 7-phosphate et phosphate.

Ces faits forment un modèle du métabolisme de la levure, qui comprend des connaissances sur les gènes, les enzymes et les métabolites (de petites molécules

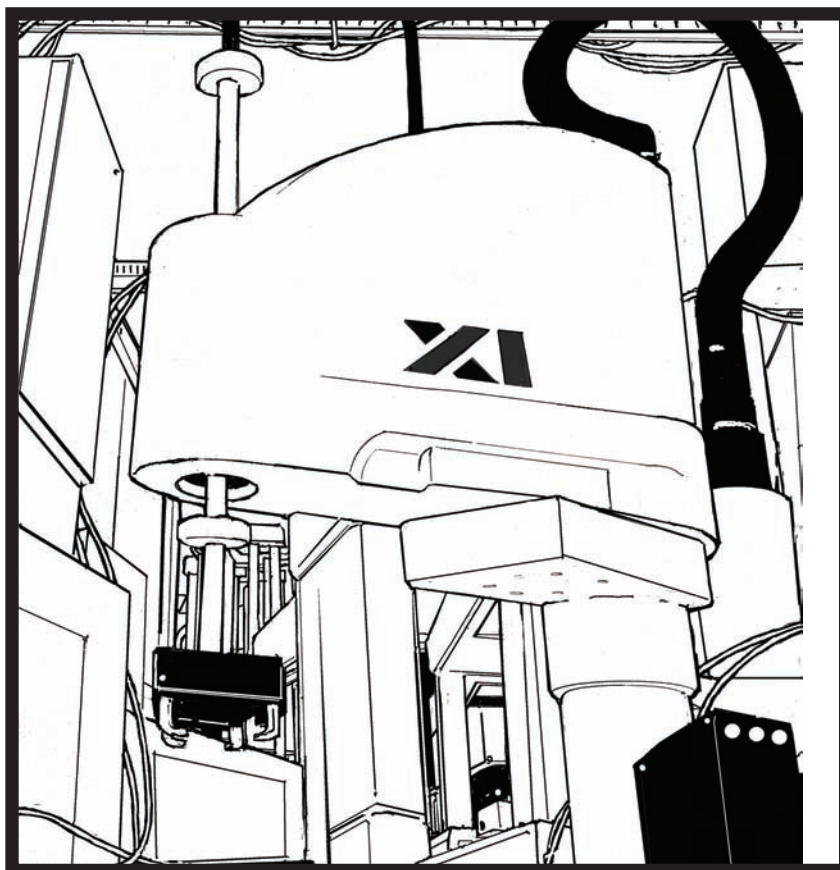
L'ESSENTIEL

✓ La conception et la réalisation de certaines expériences scientifiques exigent beaucoup de travail humain.

✓ Un robot prototype, nommé Adam, émet des hypothèses sur les gènes de levure et leurs fonctions, conçoit des expériences et les réalise.

✓ Avec l'intelligence artificielle et l'équipement robotique, Adam a découvert trois gènes particuliers de levure.

✓ L'intervention de l'homme est nécessaire au fonctionnement d'Adam, mais, ensemble, les hommes et les machines obtiendraient davantage de résultats.



produites par le métabolisme). À l'inverse d'une encyclopédie, ce modèle est converti en un logiciel qui peut agir sur les données pour faire des prédictions. Un robot scientifique peut alors combiner des modèles scientifiques abstraits avec la robotique de laboratoire pour tester et améliorer automatiquement les modèles.

Quand les chercheurs travaillent, ils formulent des hypothèses, puis les testent expérimentalement. De même, Adam émet des hypothèses nouvelles concernant la biologie de la levure, puis il en déduit les conséquences avec son modèle de métabolisme. Il réalise ensuite des expériences pour voir si les hypothèses concordent avec les observations.

D'abord, Adam a prédit quels gènes pouvaient coder les enzymes. La formulation de telles hypothèses met en jeu différentes capacités de raisonnement (voir l'encadré page 72). Pour ne retenir que les hypothèses les plus vraisemblables, Adam a utilisé ses connaissances. Par exemple, il connaissait une enzyme, la 2-aminoadipate transaminase, dont on ignorait le gène. Cette enzyme catalyse la réaction du 2-oxoadipate avec le L-glutamate pour donner le L-2-aminoadipate et le 2-oxoglutarate (la réaction inverse peut aussi avoir lieu). Ce composé participe à la multiplication des levures ; cette réaction est importante, car elle est une cible potentielle des médicaments antifongiques.

Poser des hypothèses

Pour formuler une hypothèse, à savoir quel gène de la levure pourrait coder cette enzyme, Adam a interrogé sa base de données pour trouver des gènes d'autres organismes qui codent cette enzyme. Ainsi, chez *Rattus norvegicus* (le rat brun ou rat d'égout), un gène nommé *Aadat* code l'enzyme.

Adam a examiné la séquence d'acides aminés de l'enzyme codée par le gène *Aadat* et a cherché une séquence similaire codée par le génome de la levure. Adam sait que si des séquences d'acides aminés se ressemblent, il est raisonnable d'en déduire qu'elles sont « homologues », c'est-à-dire qu'elles partagent un ancêtre commun. Adam sait aussi que si des séquences sont homologues, alors la fonction de leur ancêtre commun a pu être conservée. Par conséquent, à partir de séquences d'acides aminés similaires, Adam peut conclure que les gènes

correspondants peuvent avoir la même fonction. Adam a trouvé trois gènes de levure ayant des séquences semblables à celle d'*Aadat* : *YER152c*, *YJL060w* et *YJL202w*. Il a supposé que chaque gène code l'enzyme 2-aminoadipate transaminase.

Pour tester ses hypothèses, Adam a utilisé les souches de levure de sa collection, conservées dans son congélateur, où chaque souche a un gène qui lui a été retiré. Le robot a examiné dans des conditions particulières la croissance de trois souches de levure, auxquelles il manquait

LE PROCESSUS PERMETTANT À ADAM de faire des hypothèses et de confirmer expérimentalement les connaissances nouvelles ne nécessite pas d'intervention humaine.

respectivement les gènes *YER152c*, *YJL060w* et *YJL202w* ; il les a fait pousser avec des substances chimiques qui participent à la réaction catalysée par l'enzyme.

L'étape suivante serait de faire de multiples expériences avec chaque souche. Les financements de la recherche sont toujours limités, et les scientifiques font souvent la course pour être les premiers à résoudre un problème. Nous avons donc configuré Adam de façon à ce qu'il mette au point des expériences efficaces testant les hypothèses rapidement et à moindre coût. Ainsi, Adam suppose que chaque hypothèse a une certaine probabilité d'être vraie.

Cette supposition est controversée, et certains philosophes, tel l'Autrichien Karl Popper, ont rejeté l'idée que des hypothèses puissent avoir des probabilités associées. Toutefois, la plupart des chercheurs supposent tacitement que certaines hypothèses ont plus de chances d'être vérifiées que d'autres. Par exemple, ils se rangent souvent derrière l'idée du « rasoir d'Occam » : toutes choses égales par ailleurs, une hypothèse simple est plus probable qu'une hypothèse plus complexe. Adam prend également en compte le coût des expériences, que nous définissons comme le prix des produits chimiques utilisés. Une approche plus exacte considérerait aussi le « coût » du temps nécessaire.

Étant donné un ensemble d'hypothèses associées à des probabilités et un ensemble d'expériences ayant un coût, le but que nous avons fixé à Adam est de choisir une série d'expériences qui minimise le coût et élimine toutes les hypothèses

excepté une. Suivre cette approche de façon optimale est difficile à programmer, mais nos analyses ont montré que le choix réalisé par Adam sélectionne des expériences qui résolvent les problèmes plus rapidement et pour un meilleur coût que les autres choix possibles ; par exemple, il met en œuvre l'expérience la moins chère. Dans certains cas, Adam peut concevoir une seule expérience qui vérifie de nombreuses hypothèses. Les êtres humains ont du mal à en faire autant : ils ne considèrent en général qu'une hypothèse à la fois.

Quand le système d'intelligence artificielle d'Adam a déterminé les expériences les plus prometteuses, il utilise ses équipements robotiques pour les réaliser et observer les résultats. Adam n'est pas capable de voir directement les gènes

ou les enzymes ; il mesure quelle quantité de lumière traverse les cultures de levure. Avec ces données, Adam met en œuvre une chaîne de raisonnements compliquée qui lui permet de savoir si oui ou non les indices s'accordent avec les hypothèses sur les gènes et les enzymes. Ces chaînes de raisonnement existent souvent en science ; par exemple, les astronomes déduisent ce qui se passe dans des galaxies lointaines à partir du rayonnement qu'ils observent avec leurs instruments.

Faire des découvertes

Adam a produit et confirmé expérimentalement 20 hypothèses suggérant quels gènes codent des enzymes de la levure. Comme toute découverte scientifique, celles d'Adam devaient être confirmées. Nous avons donc vérifié les conclusions d'Adam avec d'autres sources d'information qui ne lui étaient pas accessibles et de nouvelles expériences que nous avons réalisées nous-mêmes. Sept des conclusions d'Adam étaient connues, une semblait fausse, et 12 constituaient des nouveautés scientifiques.

En vérifiant ces résultats, nous avons confirmé que trois gènes (*YER152c*, *YJL060w* et *YJL202w*) codent l'enzyme 2-aminoadipate transaminase. Pourquoi nul n'avait découvert la fonction de ces gènes auparavant ? Parce que ces trois gènes codent la même enzyme, et celle-ci peut catalyser différentes réactions ; nous n'avions pas affaire à une simple correspondance entre un gène et une fonction enzymatique (le scénario classique). Les expériences et