



Deux intelligences artificielles se mettent à dialoguer

Une équipe de l'UNIGE a conçu une IA capable d'apprendre une tâche sur la seule base de consignes verbales. Et d'en faire de même avec une IA «sœur».

Exécuter une tâche inédite sur la seule base de consignes verbales ou écrites. Puis, la décrire pour qu'une autre personne puisse la reproduire. Cette double faculté est une pierre angulaire de la communication humaine, qui résiste encore à l'intelligence artificielle (IA). Une équipe de l'Université de Genève (UNIGE) est parvenue à modéliser un réseau de neurones artificiels capable de réaliser cet exploit cognitif. Après avoir appris et accompli une série de tâches basiques, cette IA a été en mesure d'en fournir une description linguistique à une IA «sœur», qui les a exécutées à son tour. Ces résultats prometteurs, notamment pour le secteur de la robotique, sont à découvrir dans *Nature Neuroscience*.

Accomplir une tâche inédite sans entraînement préalable, sur la seule base de consignes verbales ou écrites, est une faculté propre à l'être humain. Une fois la tâche apprise, nous sommes également capables de la décrire pour qu'une autre personne puisse la reproduire. Cette double capacité nous distingue des autres espèces qui, pour apprendre une telle tâche, ont besoin de nombreux essais accompagnés de signaux de renforcement positif ou négatif, sans être capables de la communiquer à leurs congénères.

Un sous-domaine de l'intelligence artificielle (IA) - le traitement du langage naturel - cherche à recréer cette faculté humaine, avec des machines qui comprennent et répondent à des données vocales ou textuelles. Cette technique s'appuie sur des réseaux de neurones artificiels, inspirés de nos neurones biologiques et de la manière dont ils se transmettent des signaux électriques dans notre cerveau. Or, les calculs neuronaux qui pourraient permettre d'accomplir l'exploit cognitif décrit plus haut demeurent encore mal compris.

«Actuellement, les agents conversationnels utilisant l'IA sont capables d'intégrer des informations linguistiques pour produire du texte ou une image. Mais, à notre connaissance, ils ne sont pas encore capables de traduire une consigne verbale ou écrite en une action sensorimotrice, et encore moins de l'expliquer par la suite à une autre intelligence artificielle pour qu'elle la reproduise», indique Alexandre Pouget, professeur ordinaire au Département des neurosciences fondamentales de la Faculté de médecine de l'UNIGE.

Un cerveau modèle réduit

Le chercheur et son équipe ont réussi à mettre au point un modèle neuronal artificiel doté de cette double capacité, en ayant toutefois été préalablement entraîné. «Nous sommes partis d'un modèle de neurones artificiels existant, S-Bert, qui compte 300 millions de neurones et qui est pré-entraîné à la compréhension du langage. Nous

l'avons "branché" à un autre réseau, plus simple, de quelques milliers de neurones», explique Reidar Riveland, doctorant au Département des neurosciences fondamentales de la Faculté de médecine de l'UNIGE, et premier auteur de l'étude.

Lors de la première étape de l'expérience, les neuroscientifiques ont entraîné ce réseau à simuler l'aire de Wernicke, la partie du cerveau qui nous permet de percevoir et d'interpréter le langage. Lors de la seconde étape, le réseau a été entraîné à reproduire l'aire de Broca qui, sous l'influence de l'aire de Wernicke, se charge de la production et de l'articulation des mots. L'ensemble du processus a été réalisé sur des ordinateurs portables classiques. Des consignes écrites, en anglais, ont ensuite été transmises à l'IA.

Par exemple: pointer l'endroit – à gauche ou à droite – où un stimulus est perçu; répondre dans la direction opposée d'un stimulus; ou plus complexe, entre deux stimuli visuels avec une légère différence de contraste, montrer le plus lumineux. Les scientifiques évaluaient les résultats de leur modèle, qui simule l'intention de bouger, ou ici de pointer. «Une fois ces tâches apprises, le réseau a été capable de les décrire à un second réseau – une copie du premier – pour qu'il puisse les reproduire. À notre connaissance, faire dialoguer deux IA de manière purement linguistique est une première», se réjouit Alexandre Pouget, qui a dirigé ces travaux.

Pour de futurs humanoïdes

Cette modélisation ouvre de nouveaux horizons pour comprendre l'interaction entre le langage et les comportements. Elle est notamment prometteuse pour le secteur de la robotique, où le développement de technologies permettant de faire dialoguer les machines est un enjeu clé. «Le réseau que nous avons mis au point est de taille très réduite. Rien n'empêche désormais de développer, sur cette base, des réseaux beaucoup plus complexes qui seraient intégrés à des robots humanoïdes capables de nous comprendre mais aussi de se comprendre entre eux», concluent les deux chercheurs.

contact

Alexandre Pouget

Professeur ordinaire
Département des neurosciences
fondamentales
Faculté de médecine
UNIGE
+41 22 379 46 99
Alexandre.Pouget@unige.ch

Reidar Riveland

Doctorant
Département des neurosciences
fondamentales
Faculté de médecine
UNIGE
+41 22 379 56 41
Reidar.Riveland@unige.ch

DOI: [10.1038/s41593-024-01607-5](https://doi.org/10.1038/s41593-024-01607-5)

UNIVERSITÉ DE GENÈVE
Service de communication

24 rue du Général-Dufour
CH-1211 Genève 4

Tél. +41 22 379 77 17

media@unige.ch

www.unige.ch