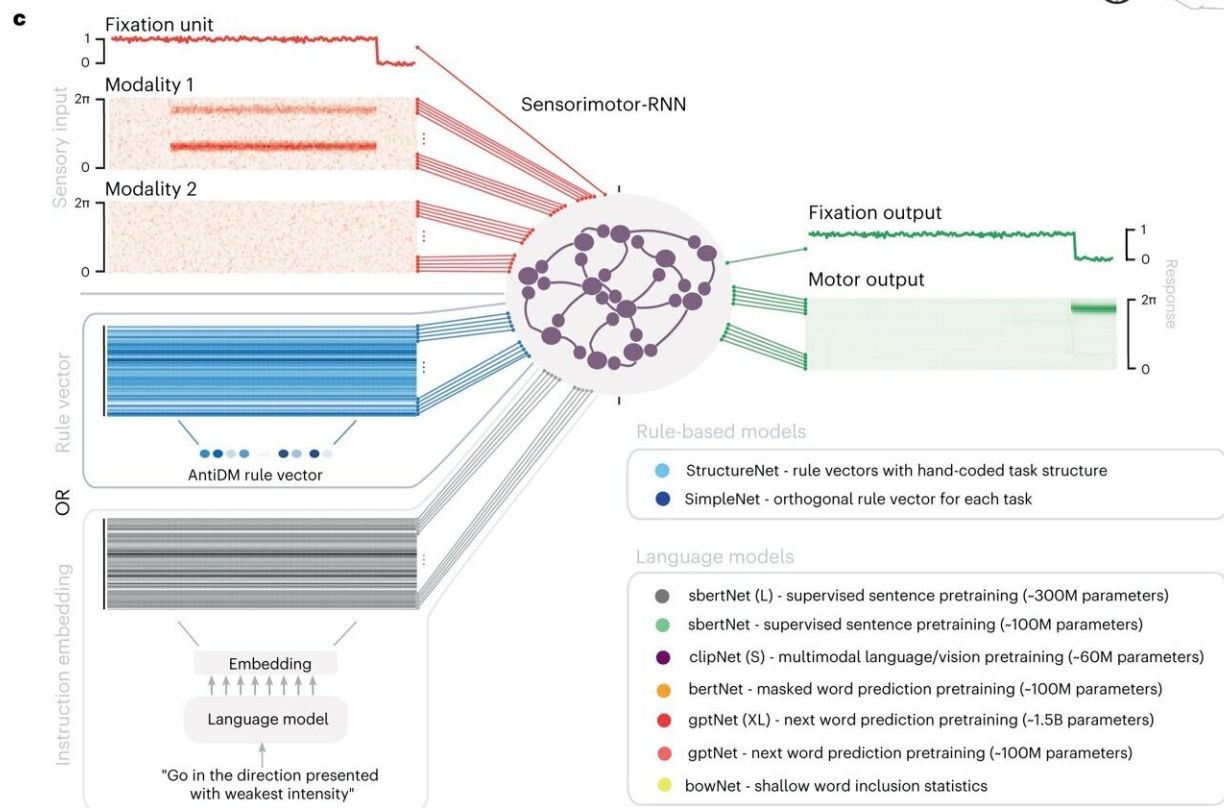# Two artificial intelligences talk to each other

March 18 2024

Tasks and models. Credit: *Nature Neuroscience* (2024). DOI: 10.1038/s41593-024-01607-5

Performing a new task based solely on verbal or written instructions, and then describing it to others so that they can reproduce it, is a cornerstone of human communication that still resists artificial intelligence (AI).

A team from the University of Geneva (UNIGE) has succeeded in modeling an [artificial neural network](#) capable of this cognitive prowess. After learning and performing a series of basic tasks, this AI was able to provide a linguistic description of them to a "sister" AI, which in turn performed them. These promising results, especially for robotics, are [published](#) in *Nature Neuroscience*.

Performing a new [task](#) without prior training, on the sole basis of verbal or written instructions, is a unique human ability. What's more, once we have learned the task, we are able to describe it so that another person can reproduce it. This dual capacity distinguishes us from other species which, to learn a new task, need numerous trials accompanied by positive or negative reinforcement signals, without being able to communicate it to their congeners.

A sub-field of [artificial intelligence](#) (AI)—Natural language processing—seeks to recreate this human faculty, with machines that understand and respond to vocal or textual data. This technique is based on artificial neural networks, inspired by our biological neurons and by the way they transmit electrical signals to one another in the brain. However, the neural calculations that would make it possible to achieve the cognitive feat described above are still poorly understood.

"Currently, conversational agents using AI are capable of integrating linguistic information to produce text or an image. But, as far as we know, they are not yet capable of translating a verbal or written instruction into a sensorimotor action, and even less explaining it to another artificial intelligence so that it can reproduce it," explains Alexandre Pouget, full professor in the Department of Basic Neurosciences at the UNIGE Faculty of Medicine.
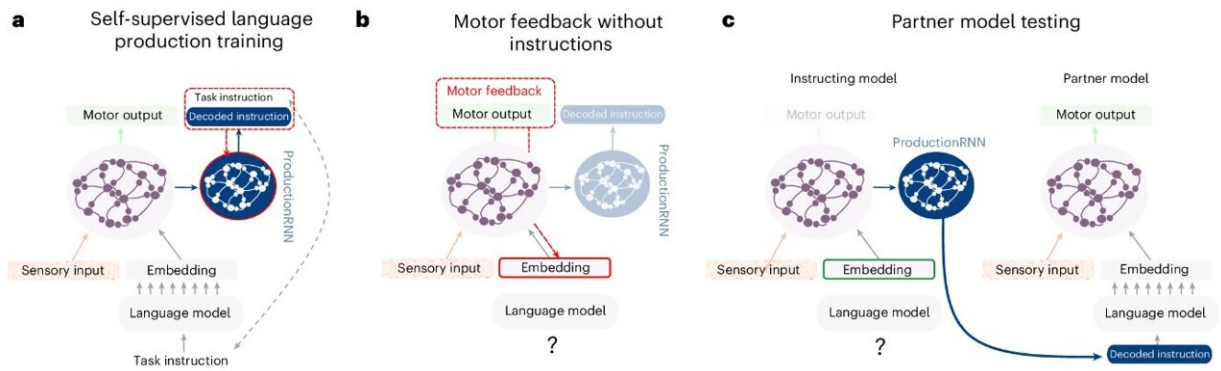
a, Illustration of self-supervised training procedure for the language production network (blue). The red dashed line indicates gradient flow. b, Illustration of motor feedback used to drive task performance in the absence of linguistic instructions. c, Illustration of the partner model evaluation procedure used to evaluate the quality of instructions generated from the instructing model. Credit: *Nature Neuroscience* (2024). DOI: 10.1038/s41593-024-01607-5

A model brain

The researcher and his team have succeeded in developing an artificial neuronal model with this dual capacity, albeit with prior training. "We started with an existing model of artificial neurons, S-Bert, which has 300 million neurons and is pre-trained to understand language. We 'connected' it to another, simpler network of a few thousand neurons," explains Reidar Riveland, a Ph.D. student in the Department of Basic Neurosciences at the UNIGE Faculty of Medicine, and first author of the study.

In the first stage of the experiment, the neuroscientists trained this network to simulate Wernicke's area, the part of our brain that enables us to perceive and interpret language. In the second stage, the network was trained to reproduce Broca's area, which, under the influence of

Wernicke's area, is responsible for producing and articulating words. The entire process was carried out on conventional laptop computers. Written instructions in English were then transmitted to the AI.

For example: pointing to the location—left or right—where a stimulus is perceived; responding in the opposite direction of a stimulus; or, more complex, between two visual stimuli with a slight difference in contrast, showing the brighter one. The scientists then evaluated the results of the model, which simulated the intention of moving, or in this case pointing.

"Once these tasks had been learned, the network was able to describe them to a second network—a copy of the first—so that it could reproduce them. To our knowledge, this is the first time that two AIs have been able to talk to each other in a purely linguistic way," says Alexandre Pouget, who led the research.

For future humanoids

This model opens new horizons for understanding the interaction between language and behavior. It is particularly promising for the robotics sector, where the development of technologies that enable machines to talk to each other is a key issue.

"The network we have developed is very small. Nothing now stands in the way of developing, on this basis, much more complex networks that would be integrated into humanoid robots capable of understanding us but also of understanding each other," conclude the two researchers.