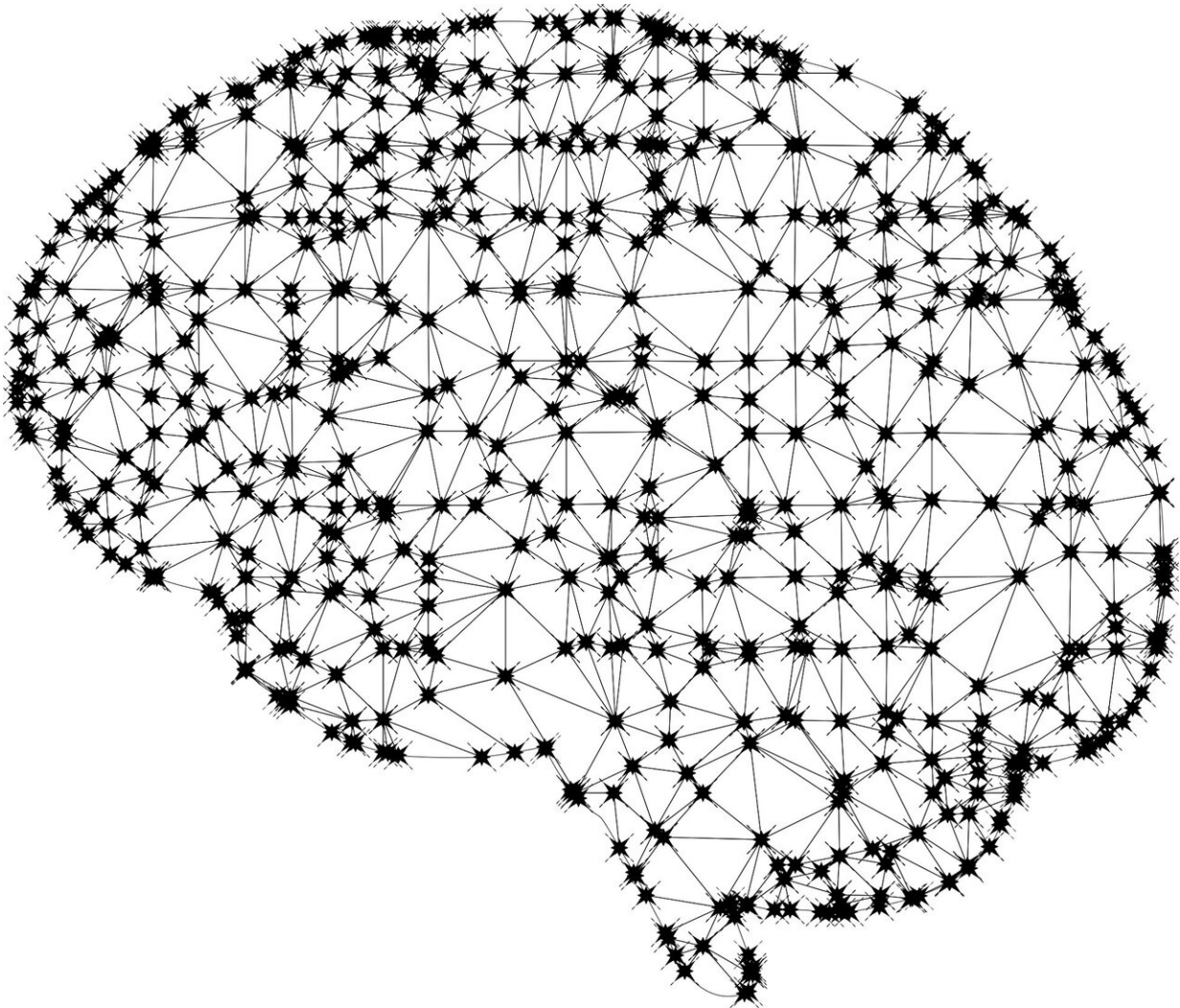# Researchers reveal roadmap for AI innovation in brain and language learning

March 19 2024, by Selena Langner

Credit: CC0 Public Domain

One of the hallmarks of humanity is language, but now, powerful new artificial intelligence tools also compose poetry, write songs, and have extensive conversations with human users. Tools like ChatGPT and Gemini are widely available at the tap of a button—but just how smart are these AIs?

A new multidisciplinary research effort co-led by Anna (Anya) Ivanova, assistant professor in the School of Psychology at Georgia Tech, alongside Kyle Mahowald, an assistant professor in the Department of Linguistics at the University of Texas at Austin, is working to uncover just that.

Their results could lead to innovative AIs that are more similar to the human brain than ever before—and also help neuroscientists and psychologists who are unearthing the secrets of our own minds.

The study, "Dissociating Language and Thought in Large Language Models," is published in *Trends in Cognitive Sciences*. An earlier preprint of the paper was released in January 2023. The research team has continued to refine the research for this final journal publication.

"ChatGPT became available while we were finalizing the preprint," Ivanova explains. "Over the past year, we've had an opportunity to update our arguments in light of this newer generation of models, now including ChatGPT."

Form versus function

The study focuses on large language models (LLMs), which include AIs like ChatGPT. LLMs are text prediction models, and create writing by predicting which word comes next in a sentence—just as a cell phone or email service like Gmail might suggest which word you might want to write next. However, while this type of language learning is extremely

effective at creating coherent sentences, that doesn't necessarily signify intelligence.

Ivanova's team argues that formal competence—creating a well-structured, grammatically correct sentence—should be differentiated from functional competence—answering the right question, communicating the correct information, or appropriately communicating. The team also found that while LLMs trained on text prediction are often very good at formal skills, they still struggle with functional skills.

"We humans have the tendency to conflate language and thought," Ivanova says. "I think that's an important thing to keep in mind as we're trying to figure out what these models are capable of, because using that ability to be good at language, to be good at formal competence, leads many people to assume that AIs are also good at thinking—even when that's not the case.

"It's a heuristic that we developed when interacting with other humans over thousands of years of evolution, but now in some respects, that heuristic is broken," Ivanova explains.

The distinction between formal and functional competence is also vital in rigorously testing an AI's capabilities, Ivanova adds. Evaluations often don't distinguish formal and functional competence, making it difficult to assess what factors are determining a model's success or failure. The need to develop distinct tests is one of the team's more widely accepted findings, and one that some researchers in the field have already begun to implement.

Creating a modular system

While the human tendency to conflate functional and formal

competence may have hindered understanding of LLMs in the past, our human brains could also be the key to unlocking more powerful AIs.

Leveraging the tools of cognitive neuroscience while a postdoctoral associate at Massachusetts Institute of Technology (MIT), Ivanova and her team studied brain activity in neurotypical individuals via fMRI, and used behavioral assessments of individuals with brain damage to test the causal role of brain regions in language and cognition—both conducting new research and drawing on previous studies. The team's results showed that human brains use different regions for functional and formal competence, further supporting this distinction in AIs.

"Our research shows that in the brain, there is a language processing module and separate modules for reasoning," Ivanova says. This modularity could also serve as a blueprint for how to develop future AIs.

"Building on insights from human brains—where the language processing system is sharply distinct from the systems that support our ability to think—we argue that the language-thought distinction is conceptually important for thinking about, evaluating, and improving large language models, especially given recent efforts to imbue these models with human-like intelligence," says Ivanova's former advisor and study co-author Evelina Fedorenko, a professor of brain and cognitive sciences at MIT and a member of the McGovern Institute for Brain Research.

Developing AIs in the pattern of the human brain could help create more powerful systems—while also helping them dovetail more naturally with human users. "Generally, differences in a mechanism's internal structure affect behavior," Ivanova says. "Building a system that has a broad macroscopic organization similar to that of the human brain could help ensure that it might be more aligned with humans down the road."

In the rapidly developing world of AI, these systems are ripe for experimentation. After the team's preprint was published, OpenAI announced their intention to add plug-ins to their GPT models.

"That plug-in system is actually very similar to what we suggest," Ivanova adds. "It takes a modularity approach where the language model can be an interface to another specialized module within a system."

While the OpenAI plug-in system will include features like booking flights and ordering food, rather than cognitively inspired features, it demonstrates that "the approach has a lot of potential," Ivanova says.

The future of AI—and what it can tell us about ourselves

While our own brains might be the key to unlocking better, more powerful AIs, these AIs might also help us better understand ourselves. "When researchers try to study the brain and cognition, it's often useful to have some smaller system where you can actually go in and poke around and see what's going on before you get to the immense complexity," Ivanova explains.

However, since human language is unique, model or animal systems are more difficult to relate. That's where LLMs come in.

"There are lots of surprising similarities between how one would approach the study of the brain and the study of an artificial neural network" like a large language model, she adds. "They are both information processing systems that have biological or artificial neurons to perform computations."

In many ways, the human brain is still a black box, but openly available

AIs offer a unique opportunity to see the synthetic system's inner workings and modify variables, and explore these corresponding systems like never before.

"It's a really wonderful model that we have a lot of control over," Ivanova says. "Neural networks—they are amazing."

Along with Anna (Anya) Ivanova, Kyle Mahowald, and Evelina Fedorenko, the research team also includes Idan Blank (University of California, Los Angeles), as well as Nancy Kanwisher and Joshua Tenenbaum (Massachusetts Institute of Technology).

Provided by Georgia Institute of Technology