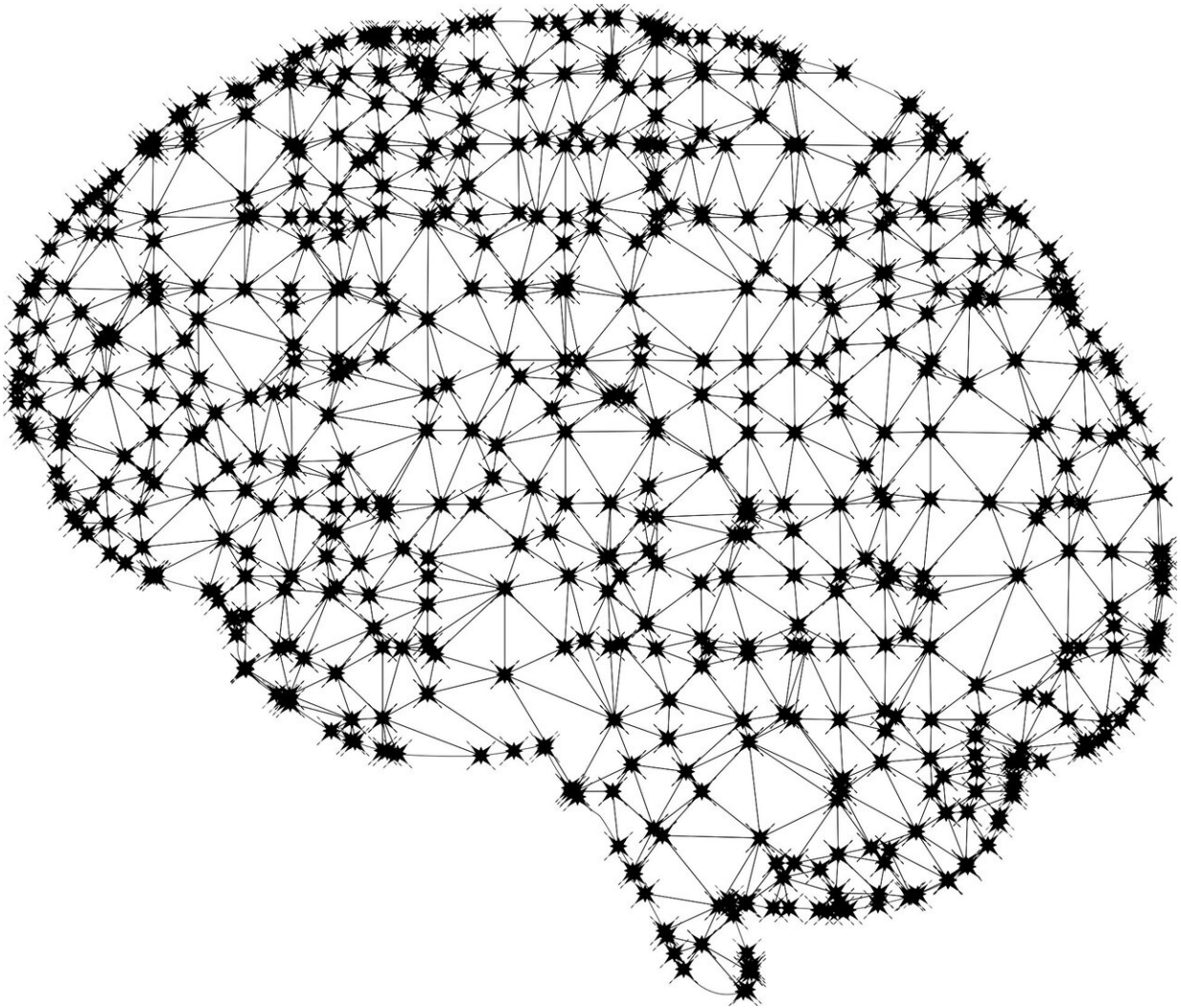


Designing AI that better understands humans' goals

December 5 2019



Credit: CC0 Public Domain

When researchers design machine learning systems, their goal is typically to automate certain functions. Instead of being fully autonomous, however, most of these systems work together with humans. In order to be truly helpful, they need to understand what goals people have.

Researchers at the Finnish Center for Artificial Intelligence (FCAI) have now taken important steps towards designing AI that understands people.

At first, the researchers taught the AI to build a model of its user—human or machine. Then, they taught it to adapt this model by following the user's actions. In practice, the researchers developed machine learning methods which combine statistics with computation, and then tested the methods in practice and in simulations. They tested the algorithms in simple situations in order to make sure they understand what exactly happens in those situations and report about the events accurately.

In the first experiment, they designed an AI teacher for the learning AI.

"This was difficult especially because the learning AI could decide what it wanted to learn," explains Samuel Kaski, the director of FCAI and professor at Aalto University. The researchers noticed that the AI learner achieved better learning results when the teacher understood what information the learner had already learned and adapted its teaching material to suit this particular learner.

In the second experiment, human users were asked to find a particular target word by using an AI-based word-search engine. The engine presents the user one word at a time, and the user then tells it whether the presented word is useful in finding the target word. If the user is looking for the word "football," for instance, they are likely to say that the first presented sport-related word is useful, if all the previous words

have been related to food.

The results of this experiment showed that the AI could help the users in finding the target words faster if it understood that, by responding to the presented words in a certain way, the user wants to direct the AI towards the right words. In other words, the AI took into account the fact that the user is trying to teach it.

According to Professor Kaski, this topic is important, as the interaction between user and AI becomes much easier when the AI understands its user's goals. "Then the human user does not need to explain in detail anymore what they expect from the AI helper."

One of the main goals of FCAI is to develop AI that understands humans and is understandable. "So far, we can build AI systems that understand the users' goals only in very simple situations, which means that designing truly helpful AI assistants calls for a lot of additional work," Kaski says.

More information: Machine Teaching of Active Sequential Learners. [aaltopml.github.io/machine-tea ... sequential-learners/](https://aaltopml.github.io/machine-tea...sequential-learners/)

Provided by Aalto University

Citation: Designing AI that better understands humans' goals (2019, December 5) retrieved 5 May 2024 from <https://techxplore.com/news/2019-12-ai-humans-goals.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.
