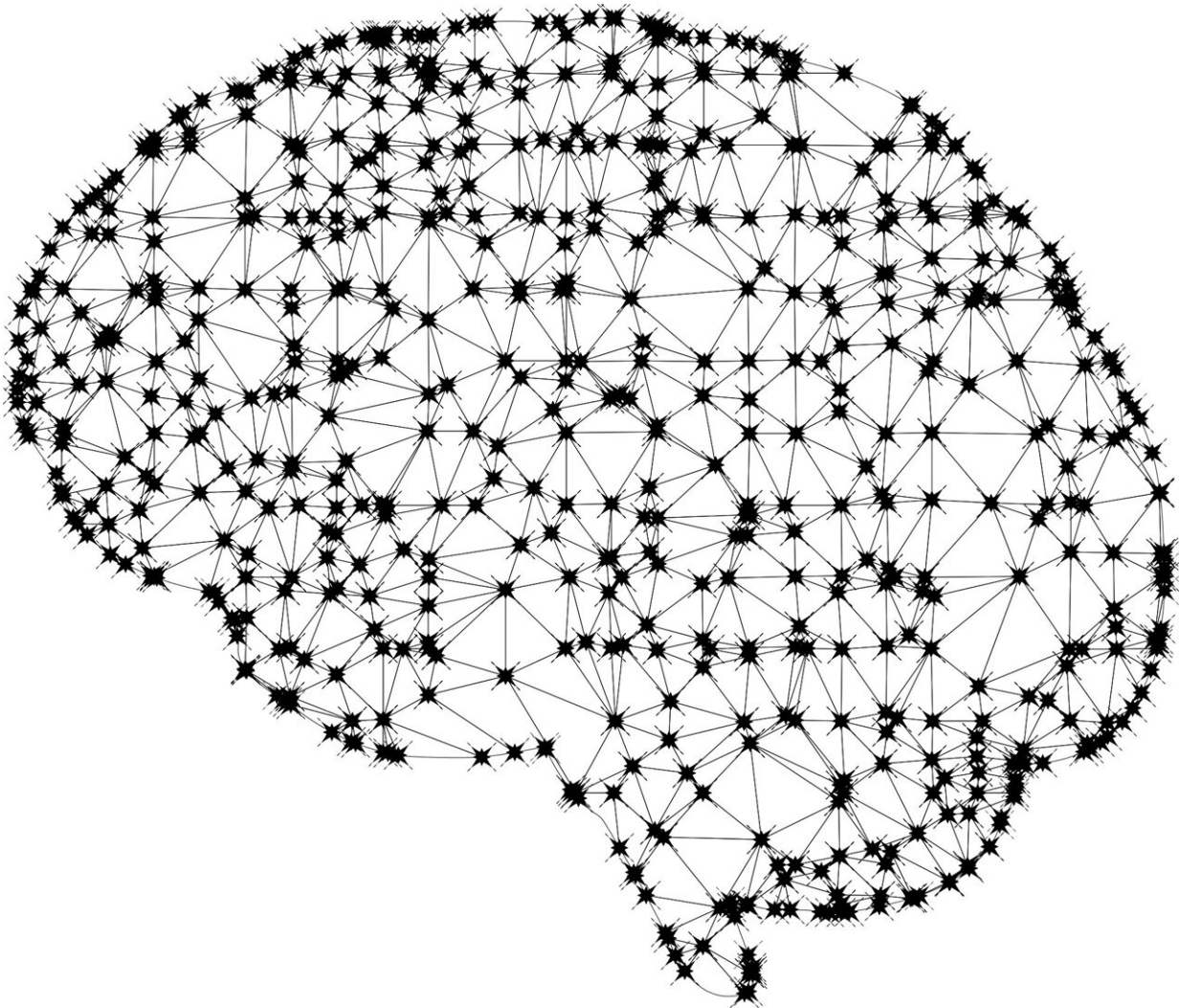# Turning senses into media: Can we teach artificial intelligence to perceive?

June 23 2022

Credit: Pixabay/CC0 Public Domain

Humans perceive the world through different senses: we see, feel, hear, taste and smell. The different senses with which we perceive are multiple channels of information, also known as multimodal. Does this mean that what we perceive can be seen as multimedia?

Xue Wang, Ph.D. Candidate at LIACS, translates perception into multimedia and uses Artificial Intelligence (AI) to extract information from multimodal processes, similar to how the brain processes information. In her research she has tested learning processes of AI in four different ways.

Putting words into vectors

First, Xue looked into word-embedded learning: the translation of words into vectors. A vector is a quantity with two properties, namely a direction and a magnitude. Specifically, this part deals with how the classification of information can be improved. Xue proposed the use of a new AI model that links words to images, making it easier to classify words. While testing the model, an observer could interfere if the AI did something wrong. The research shows that this model performs better than a previously used model.

Looking at sub-categories

A second focus of the research are images accompanied by other information. For this topic Xue observed the potential of labeling sub-categories, also known as fine-grained labeling. She used a specific AI model to make it easier to categorize images with little [text](#) around it. It merges coarse labels, which are general categories, with fine-grained labels, the sub-categories. The approach is effective and helpful in structuring easy and difficult categorizations.

Finding relations between images and text

Thirdly, Xue researched image and text association. A problem with this topic is that the transformation of this information is not linear, which means that it can be difficult to measure. Xue found a potential solution for this problem: she used kernel-based transformation. Kernel stands for a specific class of algorithms in [machine learning](#). With the used model, it is now possible for AI to see the relationship of meaning between images and text.

Finding contrast in images and text

Lastly, Xue focused on images accompanied by text. In this part AI had to look at contrasts between words and images. The AI [model](#) did a task called phrase grounding, which is the linking of nouns in image captions to parts of the image. There was no observer that could interfere in this task. The research showed that AI can link image regions to nouns with an average accuracy for this field of research.

The perception of artificial intelligence

This research offers a great contribution to the field of multimedia information: we see that AI can classify words, categorize images and link [images](#) to text. Further research can make use of the methods proposed by Xue and will hopefully lead to even better insights in the multimedia perception of AI.

Provided by Leiden University

intelligence.html