“AI doesn't really have a clue what it's doing”

AI is transforming science and does not stop at neuroscience. In this interview, we speak to Gabriele Lohmann about her research at the intersection of the brain and algorithms. She talks about the use of artificial intelligence as a research tool, the decoding of brain activity, and why large language models do not understand what they are doing.

The fields of neuroscience and artificial intelligence (AI) are closely intertwined. In what ways do these two fields benefit from each other?

For one thing, current AI models — artificial neural networks — are modeled on the structure of the human brain. These networks have been around since the 1950s, inspired by neurons and their connections. Naturally, there are significant differences between brains and artificial neural networks. Nevertheless, AI and neuroscience enrich each other. AI learns from the brain, and neuroscientists benefit from AI systems that support their work.

In what ways is AI currently helping with neuroscience research?

For instance, AI is used for image processing, which is a common challenge in magnetic resonance imaging (MRI). We generate brain images that need to be analyzed. Automated image processing programs have existed for around 40 years, but have never been particularly reliable. It is only in the last decade that AI research has made huge strides in this area. This is evident in Alzheimer's diagnostics, for example: To diagnose the disease, the size of the hippocampus is determined from an MRI scan. This is one of the most well-established markers of the condition because the hippocampus is one of the first brain regions to shrink in Alzheimer’s patients. Manual segmentation, that is, tracing the hippocampus by hand, is extremely time-consuming. AI tools can now perform this task reliably and in a fraction of the time.

But how can we be sure that what the AI detects actually corresponds to reality?

With MRI scans, this is easy, since the image is right in front of you; the radiologist can quickly verify the result. Here, the AI essentially just serves as a tool to speed up the process.

Doesn’t the use of AI in brain imaging raise the fundamental problem that you need very large amounts of data for training, possibly more than is available?

That does indeed happen. Here’s a current project from our research group: One of my doctoral students is working on early detection of Alzheimer’s. Many elderly people exhibit mild cognitive impairments that do not yet warrant an Alzheimer's diagnosis, but which could develop into one. Predicting this requires a highly accurate segmentation of MRI images. Fortunately, there are large databases with thousands of images from all over the world from 3-Tesla MRI machines, which are standard issue in many clinics. This provides more than enough material with which to train an AI. However, when we wanted do the same with data from our 9.4-Tesla MRI machine, we were faced with the problem that such high-field scanners are very rare: We only had scans from around 100 subjects, taken at our own institute. My doctoral student's solution was to use the model trained on 3-Tesla data as a starting point and adapt it for 9.4 Tesla. And as it turned out, this worked well.

Can AI also help researchers by extracting patterns from large amounts of data, generating its own hypotheses based on these patterns?

Yes, in a sense – for a process known as decoding: Researchers measure the brain activity of test subjects in a scanner and attempt to reconstruct their thoughts. This currently works passably well with the help of diffusion models, the technology behind deepfakes. In decoding, a person is placed in a scanner and shown images, and the AI tries to guess what the subject

sees. Based on initial training with large amounts of data, the model recognizes patterns and uses deepfake technology to 'fake' the rest.

Is this purely basic research, or could it have practical applications? Could it, for example, help patients with locked-in syndrome?

In the long term, this is certainly a possibility. But we're not there yet.

How will the use of AI in research evolve over the next five to ten years?

Right now, AI is extensively used as a tool, for example to write articles or generate suggestions. You can feed a large language model like ChatGPT a manuscript and ask questions such as, 'How important is this? What would you do next?' The suggestions are surprisingly good — it's truly amazing how intelligent they seem. However, if you look more closely, you realize that these models don't really have a clue what they're doing. Essentially, they perform pattern recognition: They process huge amounts of data and calculate, statistically, which word is likely to come next. There's no real understanding behind it. That's why I'm particularly interested in testing what large language models comprehend. Existing AI benchmarks are routinely overcome after around two years because of the rapid progress in development. I am therefore trying to create a new test that focuses on logical thinking. It will likely be quite a while before an AI masters that.

Press release

13-May-2026
Source: Max Planck Institute for Biological Cybernetics

Further information

PD Dr. Gabriele Lohmann
Group Leader
Phone: +49 (0) 7071 601 931
Email: gabriele.lohmann(at)tuebingen.mpg.de

- Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.
- Max Planck Institute for biological cybernetics