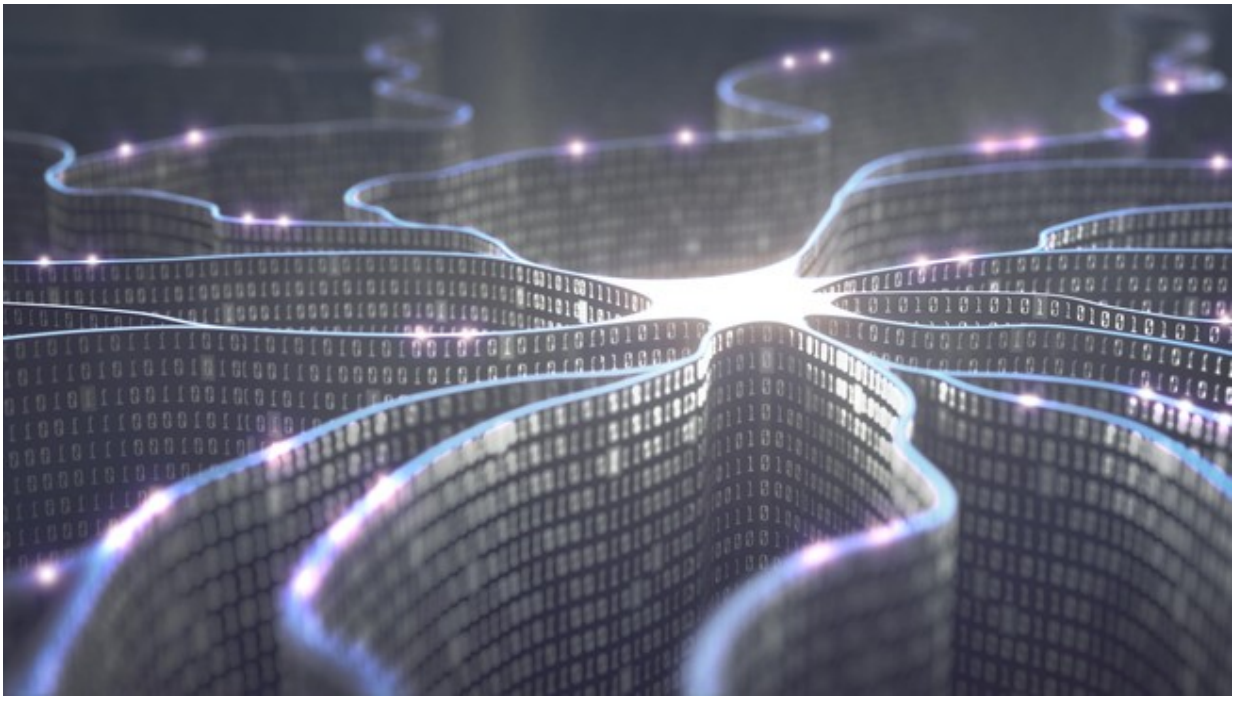
Is reliable artificial intelligence possible?

March 15 2017, by Hillary Sanctuary



Credit: ktsimage

n the quest for reliable artificial intelligence, EPFL scientist Marcel Salathé argues that AI technology should be openly available. He will be discussing the topic at this year's edition of South by South West on March 14th in Austin, Texas.

Will artificial intelligence (AI) change the nature of work? For EPFL theoretical biologist Marcel Salathé, the answer is invariably yes. To

him, a more fundamental question that needs to be addressed is who owns that artificial intelligence?

"We have to hold AI accountable, and the only way to do this is to verify it for biases and make sure there is no deliberate misinformation," says Salathé. "This is not possible if the AI is privatized."

AI is both the algorithm and the data

So what exactly is AI? It is generally regarded as "intelligence exhibited by machines". Today, it is highly task specific, specially designed to beat humans at strategic games like Chess and Go, or diagnose skin disease on par with doctors' skills.

On a practical level, AI is implemented through what scientists call "machine learning", which means using a computer to run specifically designed software that can be "trained", i.e. process data with the help of algorithms and to correctly identify certain features from that data set. Like human cognition, AI learns by trial and error. Unlike humans, however, AI can process and recall large quantities of data, giving it a tremendous advantage over us.

Crucial to AI learning, therefore, is the underlying data. For Salathé, AI is defined by both the algorithm and the data, and as such, both should be publicly available.

Deep learning algorithms can be perturbed

Last year, Salathé created an algorithm to recognize plant diseases. With more than 50,000 photos of healthy and diseased plants in the database, the algorithm uses artificial intelligence to diagnose plant diseases with the help of your smartphone. As for human disease, a recent study by a

Stanford Group on cancer showed that AI can be trained to recognize skin cancer slightly better than a group of doctors. The consequences are far-reaching: AI may one day diagnose our diseases instead of doctors. If so, will we really be able to trust its diagnosis?

These diagnostic tools use data sets of images to train and learn. But visual data sets can be perturbed that prevent deep learning algorithms from correctly classifying images. Deep neural networks are highly vulnerable to visual perturbations that are practically impossible to detect with the naked eye, yet causing the AI to misclassify images.

In future implementations of AI-assisted medical diagnostic tools, these perturbations pose a serious threat. More generally, the perturbations are real and may already be affecting the filtered information that reaches us every day. These vulnerabilities underscore the importance of certifying AI technology and monitoring its reliability.

Provided by Ecole Polytechnique Federale de Lausanne