

Why artificial intelligence is likely to take more lives

December 11 2018, by Martial Mermillod



Credit: Geralt/Pixabay

Artificial neurons for deeply intelligent machines – this is the new artificial intelligence (AI) revolution, led by <u>Geoffrey Hinton</u> and his team since 2012. That year, Hinton, an expert in cognitive science at the University of Toronto and a researcher at <u>Google Brain</u>, demonstrated



the striking effectiveness of a deep neural network (DNN) in an imagecategorisation task.

In the wake of these remarkable results, universities – and international corporations – invested massively in the promising and fascinating field of AI. Yet despite the impressive performance of DNNs in a variety of fields (visual and vocal recognition, translation, medical imagery, etc.), questions remain regarding the limits of deep learning for other uses, such as antonymous vehicles.

To understand the limits of AI in its current state, we need to understand where DNNs come from and, above all, which areas of the human brain they are modelled on – little is known about this in industrial engineering, and even in some research centres. Since the dawn of this new revolution, deep learning is sometimes used as a kind of "magic wand", with scant attention to its background or limitations. "For a meaningful artificial intelligence", the title of a recent report by French mathematician Cédric Villani, is evidence of the profound ambivalence surrounding this topic.

Where did deep learning come from?

The beginnings of artificial neural networks date back to the 1940s, with the pioneering discoveries in neuroscience and psychology of Warren McCulloch and Walter H. Pitts (who provided the <u>first mathematical</u> <u>model of a neuron</u>) and <u>Donald Hebb</u> (who described the mechanisms of synaptic learning). These researchers wanted to understand how neurons, the basic building blocks of the brain, could generate the psyche.

Their seminal work led to the creation of the first artificial neuronal network, the <u>Perceptron</u>, designed in 1958 by American psychologist Frank Rosenblatt. Naturally, initial research was followed by significant developments based, for instance, on the neuroscientific studies of <u>Alan</u>



L. Hodgkin and Andrew F. Huxley describing the temporal dynamics of neural integration, and on research in computer science and mathematics by <u>Bernard Widrow</u> and <u>Ted Hoff</u>, who suggested the use of stochastic gradient descent algorithms as a more effective way to modify the synaptic connections in neural networks.

These mathematical optimisations were further developed in the 1980s with research in <u>cognitive science</u> by David Rumelhart, Geoffrey Hinton and James McClelland, members of the <u>Parallel Distributed Processing</u> <u>Research Group</u>. Their work helped optimize the modification of synaptic connections in deep neuronal layers and led to the creation of the <u>Multilayer Perceptron</u> (MLP). DNNs, developed by researchers such as Geoffrey Hinton, Yann LeCun and Yoshua Bengio, are its direct descendants.

Is there a link between deep learning and the brain?

Although DNNs were originally developed through interdisciplinary work and inspired by brain function, one might wonder to what extent these algorithms still constitute a simulation of the human brain. They were designed to carry out such tasks as image recognition and categorisation. In order to do this, DNNs use various <u>convolutional</u> and <u>pooling</u> layers prior to image recognition.

With regards to convolutional layers, the work of David Hubel and Torsten Wiesel in the 1960s, and Leonie Jones and Derecke Palmer in the 1980s, demonstrate the usefulness of this method in simulating the likely response of neurons in the primary visual cortex. Several studies in cognitive science, <u>including our own work</u>, use this process as a neuroinspired system to simulate the response of perceptual neurons in the <u>primary visual cortex</u> for instance.





An example of the research demonstrating the similarities in the way deep neural networks and the ventral stream of the visual cortex operate. Credit: Kuzovkin, Vicente, Petton, Lachaux, Baciu, Kahane et Aru, 2018., Author provided

With regards to pooling, various studies in the fields of neuroscience and cognitive psychology over the last thirty years have demonstrated how the brain carries out this process of abstraction in the ventral visual stream. The work of Rufin Vogels and Keiji Tanaka show how this stream enables visual identification and categorisation, independently of the surface properties of an image, such as texture, colour, distance, or the position of objects within the image. These brain areas are therefore sensitive to the same information as the perceptual layers learned by a DNN during the process of pooling.

Even more surprising, research by <u>Rodrigo Quian Quiroga</u> and his colleagues demonstrates the existence of specific neurons for concepts or identities (for example, a "Jennifer Aniston" neuron, or a "Tower of Pisa" neuron). These fire in response to direct exposure to a concept,



such as seeing the name "Jennifer Aniston" in print. What is more exciting for the future of AI is that the work of Mr. Quiroga demonstrates that this neuronal activity is correlated with the *conscious* perception of a stimulus in the environment.

To summarise, although they are simplified and mathematically optimised as compared to a biological brain, DNNs reproduce very similar processes as a very specific area in the cortex (namely, the occipito-temporal cortex). Using MRI or electrodes implanted in the brain, recent studies in cognitive neuroscience demonstrate similarities in the functioning of DNNs and these specific brain areas.

Is AI more reliable when inspired by the brain?

Basic interdisciplinary research on DNNs has produced impressive tangible results in a wide range of areas: visual recognition and categorisation, vocal recognition, translation, the game of go, musical composition, to name but a few. Unfortunately, through a lack of understanding of the cognitive science underpinning them, DNNs are still too often used as a kind of magic wand to solve any and all problems.





It is possible to improve a neuromorphic system's ability to anticipate by simulating the recurrent loops from associative areas to perceptual areas at work in the human brain. Credit: Mermillod, Bourrier, David, Kauffmann, Chauvin, Guyader, Dutheil et Peyrin, 2018., Author provided

To take the example of driverless cars, thoughtlessly coupling DNNs to the vehicle's control systems would be highly risky: it would be tantamount to asking a taxi driver who lost over 80% of his <u>brain</u> <u>function</u> in an accident (leaving only the visual ventral stream) to drive a car. Requiring these systems to do more than what they were originally designed for may lead to catastrophic accidents.

The human brain areas involved in anticipation (see below), spatial orientation and the sensory-motor functions required for driving in a complex environment are very different from the neural processes at



work within the visual ventral stream. Located in the occipito-parietal cortex, the neuronal processes involved in the understanding and planning are very different from those taking place in the visual ventral stream! These are very different neurons, sensitive to distance, position and speed – all fundamental parameters for determining how we behave in the environment.

Undiscriminating use of DNNs (or other artificial systems) without reference or comparison to the neuro-inspiration behind the various cognitive functions is not only ineffective but outright dangerous. We do not claim that neuro-inspiration is the only effective way toward safer AI. However, given AI's tumultuous past, and taking into account the now proven effectiveness of neuro-inspired systems like DNNs as compared to previous engineering methods (for visual recognition, for example), we believe it is essential to understand how the brain performs other cognitive functions (motor control, multisensory integration, etc.) in order to compare this to current engineering techniques for performing these functions, and produce a safer, more efficient AI.

AI research conducted in closer collaboration with the cognitive sciences would enable us to:

- understand and simulate <u>brain</u> areas which are not yet encompassed by <u>deep learning</u>.
- develop more reliable and effective AI as compared to human performance.

This challenge requires interdisciplinary research involving not only mathematics and computer science, but also neuroscience and cognitive psychology, as well as research in electronics and physics to develop the neural processor units (NPU) currently being designed. We have the opportunity to finally overcome the limits of the Turing-Von Neumann machines that have dominated electronics and information technology



since World War II.

This article is republished from <u>The Conversation</u> under a Creative Commons license. Read the <u>original article</u>.

Provided by The Conversation

Citation: Why artificial intelligence is likely to take more lives (2018, December 11) retrieved 28 April 2024 from <u>https://phys.org/news/2018-12-artificial-intelligence.html</u>

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.