Ali H. Sayed | IEEE Signal Processing Society President | ali.sayed@epfl.ch

Intelligent Machines and *Planet of the Apes*

I love the last scene from the movie *Planet of the Apes* (1968), which revealed how the human race destroyed its beautiful planet only to be overtaken by intelligent apes. In that movie, humans were the victims of their own intelligence.

Today, we live in the midst of media frenzy with almost daily reminders about the possibility of humans being overtaken by intelligent machines. Notables like the late physicist Stephen Hawking (1942–2018) and entrepreneur Elon Musk of Tesla and SpaceX fame have issued dire warnings about the existential danger posed to our way of life by the evolution of artificial intelligence (AI). Microsoft founder and philanthropist Bill Gates has been on both sides of the issue, warning about its dangers in one instance and embracing its potential in another. In my view, Gates' position is the correct one. There is immense potential in the field, with so much good that can be done for society at large, but there are also dangers and important ethical questions. Google's recent illustration in May 2018 of its digital assistant's ability to mimic the human conversational style in a very realistic manner has raised eyebrows about the danger of using these machines in deceptive practices.

Many technologies are driving the progress that we are witnessing in intelligent machines, including advances in deep learning and machine learning, image processing, speech processing, and computer vision, as well as the availability of computing power and storage capabilities necessary to process massive amounts of data. It is no exaggeration to state that the progress is mainly driven by the last two technologies, namely, advances in computational power and storage: we now have faster computing machines and almost limitless storage capabilities for vast amounts of data. The foundational theories and algorithms continue to be largely similar to what we have known for decades: backpropagation, online learning, reinforcement learning, stochastic algorithms, pattern recognition, speech and image processing techniques, etc. That is why much of the eye-catching headlines about impressive AI applications are emanating from development labs associated with large companies with great resources to spare on computing and storage such as Google and its DeepMind affiliate, Facebook, Amazon, Microsoft, IBM, and others. Examples of successful AI applications include Google Translate, Alexa from Amazon, and powerful game machines by DeepMind and IBM capable of defeating world champions on their own turf. In 2017, DeepMind's AlphaGo playing machine defeated the world's leading GO player in an impressive feat primarily using reinforcement-learning techniques. The game of GO is particularly challenging given the size of its state space with close to 10^{170} states; compare this value with the estimated number of atoms in the universe, which stands at about 10^{80}. This is, of course, no indication of the "intelligence" of the AlphaGo machine. Its performance is simply a reflection of a brute force approach to exploration and analysis.

What limits the "intelligence" of the machines we are designing today is the fact that they can only be as good as the data used to train them.

These amazing achievements provide vivid indications of what is possible. They also lend wings to our imagination and help drive up expectations, sometimes in unrealistic ways. Some of the expectations for AI are likely to remain unfulfilled for one simple reason: the supporting science is still not there (and we may not even get there). While many AI systems have proven enormously successful in a range of applications, we need to recognize that these systems continue to be fundamentally "unintelligent." What we call "artificial intelligence" is not real intelligence. And what we call "deep learning" is not really deep. The qualification "deep" in deep learning is not referring to a supersized ability to learn and think but, rather, to an architectural structure involving a multitude of layers in a neural network implementation (representing depth).

Digital Object Identifier 10.1109/MSP.2018.2834779
Date of publication: 27 June 2018

We need to be prudent about how we name our technologies because names matter, and they may end up conveying exaggerated expectations. The "intelligent" devices of today, and the accompanying "deep" learning structures, are far from competing with the cognitive abilities and abstract thinking of the human mind. Humans can learn and create abstract concepts as well as apply them across different tasks whether in solving a puzzle, playing a game of chess, writing a piece of poetry, or proving a mathematical theorem. Real abstract reasoning is beyond the reach of current AI systems. HAL 9000, the fictional artificial intelligent character in the 1968 movie *2001: A Space Odyssey,* was capable of multitasking and able to perform speech recognition, facial recognition, lip reading, and reasoning and could even understand emotions. But that was fiction then and continues to be fiction today. We regularly witness failures that are characteristic of what could go wrong with current technologies, with some serious outcomes such as self-driving cars being involved in fatal accidents in 2017, Microsoft's chatbot being easily manipulated to posting racist comments in 2016, and even the ID facial recognition system on the new iPhone X being fooled by a three-dimensional-printed face mask in 2017. These are relatively recent examples, which raise questions about the reliability of even the most up-to-date technologies.

What limits the "intelligence" of the machines we are designing today is the fact that they can only be as good as the data used to train them. These devices spot and react to patterns they have learned from the training data. If, for example, a self-driving car was never trained to recognize an individual crossing the street in the dark, it will most likely not learn how to respond to such an event or may respond in some erratic manner. An "intelligent" machine should at least have the ability to reason and make logical deductions based on the circumstances. But even then, machines would continue to be limited in their "intelligence" because of their high degree of specialization. For instance, the powerful AlphaGo game machine that defeated the world champion in the game of GO will perform miserably at another game, no matter how simple it is. This is because the machine was designed for the GO game. Likewise, an AI system designed to win chess games will perform poorly if we change some simple rules of the game such as limiting the movements of the queen piece or perhaps reversing the colors of the black and white squares. "Intelligent" machines are only good for the specific tasks they have been designed for, and they will achieve that level of "excellence" only if they have been trained well enough with a massive amount of data. Many would perceive them as "intelligent" only because they can perform their tasks much faster than normal human abilities, which is a fallacy. Certainly, a car traveling at 100 km/h is not "intelligent" because it is faster than a human. It is simply much more efficient in this particular task. Extrapolating from "being good at something" to "being intelligent about everything" is the problem we are facing today with the media frenzy around AI and the learning methods powering them. Fortunately, scholars and scientists driving the field are far more reasoned in their approaches and expectations. They build their understanding one step at a time and benefit from a fertilization of ideas across fields.

Many of the methodologies and techniques driving machine learning and deep learning today have been part of the signal processing repertoire for decades.

Consider, for example, the widely successful convolutional neural networks (CNNs), which constitute one of the main drivers in the ongoing data revolution since their use in 1989 by LeCun and his collaborators in the recognition of handwritten ZIP codes [1]. CNNs are an outgrowth of an architecture known as "neocognitron," which was developed a decade earlier by Fukushima (1980) [2]. This structure was, in turn, motivated by the discoveries in the 1960s of the corecipients of the 1981 Noble Prize in Medicine David H. Hubel (1926–2013) and Torsten N. Wiesel (1924–present) for their work on understanding the visual cortex of cats [3]. Since their launch, CNNs have revolutionized image processing and computer vision applications as well as speech processing and other fields. What I find most interesting about this story is not the many wonderful applications that have emerged since then, such as automatic photo tagging, scene understanding, and face recognition. For example, police in China have started using sunglasses connected to the Internet and equipped with facial recognition technology to spot criminal suspects in crowds. All of these applications do not impress me because signal processing experts like you and me understand well the algorithms that drive the systems and how these systems can be trained given sufficient data. What amazes me the most about the origins of the CNN approach is something completely different. I stand in awe at seeing the progress we have achieved today from simply studying the visual cortex system of the cat! Just imagine how much more progress we can achieve by studying the immense biodiversity that lives around us, including tiny flies!

Signal processing to the rescue

There are immense opportunities for signal processing to empower AI systems, simply because we are the discipline that specializes in extracting information from data and in understanding data, representing data, and projecting into the future. We are masters of information processing. Actually, many of the methodologies and techniques driving machine learning and deep learning today have been part of the signal processing repertoire for decades, including stochastic algorithms, the backpropagation algorithm, neural networks, Markov models, mixture models, Bayesian inference, classification, etc. The fundamental science driving most AI innovations is largely the same we have had from years past. What is pushing AI further today is the availability of massive data to train these machines, with the accompanying computing power necessary to carry out the processing. Much of the progress is data and computing. It is no wonder that the most successful AI applications today are in domains where

we have access to large amounts of data, such as speech, imaging, search engines, and games. We can complement this progress with a more fundamental science-based understanding of the processing abilities and limitations of data-driven designs. That is where signal processing approaches can contribute.

What is missing in much of the ongoing work on AI methods and deep-learning methods is an unequivocal understanding of the science behind how these systems operate and how they reach their decisions. Why are certain decisions preferred and in what ways are they optimal or reasonable? There is overreliance on learning by training, which, by default, biases the machines to operate within the boundaries dictated by the training data. That is one reason, for example, why there are concerns about using AI assistants in courtrooms to decide the sentences for convicted individuals. Most leading scholars and researchers in the AI field are aware of the limitations; they are thoughtful individuals with strong interest in building up their theories on strong foundations. But many of them do not control the outlets that propagate the "fake news" and the inflated expectations. Moderation, along with prudent and proven science, should be the norm.

For example, in a recent talk I attended on deep learning, I was not surprised by how many times the speaker repeated the words "doing this *plus* a few other tricks, you get this or that result." Whenever I hear the words "a few other tricks" in a scientific presentation, it raises a red flag in my mind because they convey to me that we still do not have sufficient understanding of the subject matter, and, therefore, one needs to try "this trick" or "that one" to get the system to work. These words, and the speaker's honesty, are a testament to the aforementioned fact, namely, that we still lack a full understanding of how AI systems work. Experts in signal processing will not hesitate to "use tricks" as well. However, they will still strive to understand "why the trick is needed," and the choice of "which trick to use" is often guided by some underlying theory. It is not uncommon for signal processing papers to be turned down if they lack sufficient theoretical justification. Signal processing experts can play an important role in solidifying the foundational basis for the ongoing data-driven revolution. Let me give you one example from the past. Consider the nearest-neighbor (NN) rule, which assigns feature vectors to the label of their closest neighbors. This is a simple classification rule, with an intuitive appeal and construction. However, a scholar with a signal processing mind would want to know more. We would like to understand why the rule works and what performance guarantees it has. The beautiful seminal result by Cover and Hart (1967) essentially showed that the probability of error of the NN rule is bounded by twice the probability of error of the Bayes classifier regardless of the underlying distribution [4]! In other words, as stated in their paper, "any other decision rule… can cut the probability of error by at most one half." A remarkable conclusion!

Just like signal processing is the "science behind our digital life," we also have a role to play in developing the "science behind the data-driven revolution."

While the topic of AI fascinates us all, some of its progress is driven by contributions from real signal processing experts like you and me, whether you are working on learning algorithms, natural language processing, image processing, computer vision, or data science. Just like signal processing is the "science behind our digital life," we also have a role to play in developing the "science behind the data-driven revolution." I agree that there is so much potential ahead of us, with superlative applications that will make our lives easier. At the same time, as science often does, we need to progress with reason and confidence to develop working systems away from unrealistic expectations or even catastrophic consequences. These ramifications were already foreseen by the two forward-looking movies *Planet of the Apes* and *2001: A Space Odyssey*, which were both released on the same day in April 1968! One movie looked at humans and the other at machines. In the first movie, the human intelligence led to a catastrophic end, while the second movie showed what could go wrong with AI with the "intelligent" HAL machine having to be turned down before it was too late!

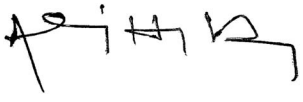
References

[1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jacke, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[2] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.

[3] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, 1962.

[4] T. M. Cover and P. E. Hart, "Nearest-neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967.



SP