R&D

AI will soon 'evolve' like humans – and civilisation as we know it will change forever

Deep-learning pioneer Jürgen Schmidhuber explains why human-level AI is near

By **JÜRGEN SCHMIDHUBER**

1 day ago

There's no reason why machines cannot be curious and creative

Credit **David Doran**

WIRED   **R&D**

When I was a teenager in the 70s, my goal was to build a self-improving AI smarter than myself, then retire. So I studied maths and computer science.

For the cover of my 1987 diploma thesis, I drew a robot that bootstraps itself in seemingly impossible fashion. The thesis was very ambitious and described the first concrete research on a self-rewriting "meta-program" which not only learns to improve its performance in some limited domain, but also learns to improve the learning algorithm itself, and the way it meta-learns the way it learns.

This was the first in a decades-spanning series of papers on algorithms for recursive self-improvement, with the goal of building a super-intelligence. I predicted that, in hindsight, the ultimate self-improver will seem so simple that students will be able to understand and implement it. I said it's the last significant thing a man can create, because all else follows from that. I am still saying it.

What kind of computational device should we use to build AIs? Physics dictates that future efficient computational hardware will look a lot like a brain-like recurrent neural network (RNN), a general-purpose computer with many processors packed in a compact volume connected by wires, to minimise communication costs. Your cortex has more than ten billion neurons, each connected to 10,000 other neurons on average. Some are input neurons that feed the rest with data (sound, vision, touch, pain, hunger). Others are output neurons

en in between, where thinking takes place.

on strengths, which determine how strongly neurons influence each other, and long experience. It's the same for our artificial RNNs.
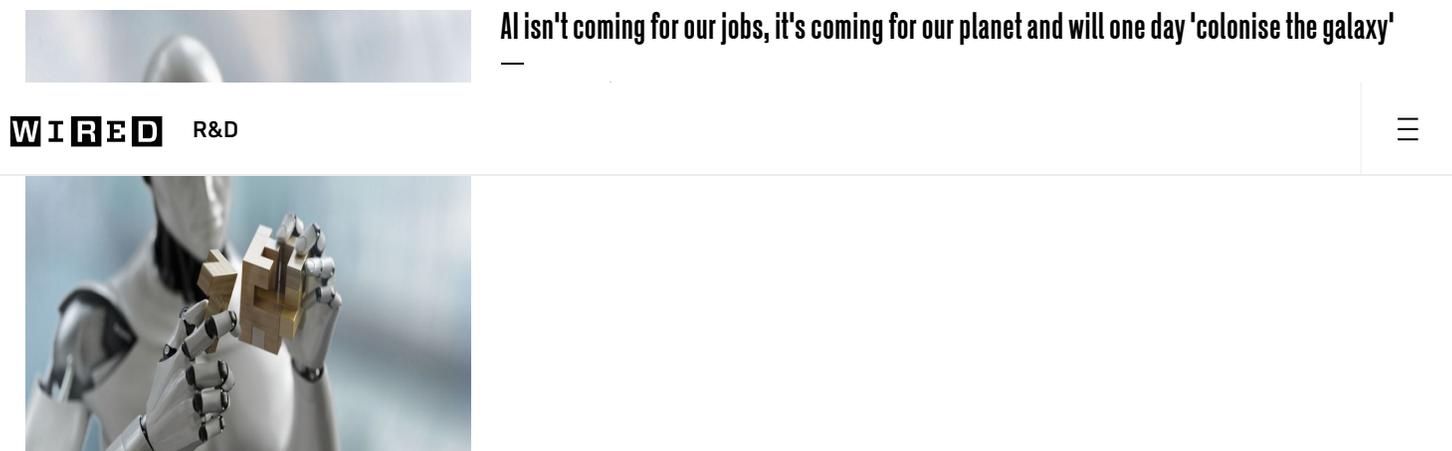
l networks (NNs) and others is that we figured out ways of making NNs deeper Ns, which have feedback connections and can, in principle, run arbitrary ng with the environment. In 1991, I published on "very deep learners"1, 3 - eight-layer nets of the Ukrainian mathematician Alexey Grigorevich earning in the 60s. By the early 90s, our RNNs could learn to solve many

Most current commercial NNs need teachers. They rely on a method called backpropagation, whose present

form was first formulated by Seppo Linnainmaa in 19703and applied to teacher-based supervised learning NNs in 1982 by Paul Werbos. However, backpropagation didn't work well for deep NNs.

In 1991, Sepp Hochreiter, my first student working on my first deep-learning project, identified the reason for this failure: the so-called vanishing gradient problem. This was then overcome by a now widely used deep learning RNN called long short-term memory (LSTM) developed in my labs since the early 90s [2,3]. In 2009, LSTM became the first RNN to win international pattern-recognition contests, through the efforts of Alex Graves, another former student. The LSTM principle has become a basis of much of what's now called deep learning.

When people ask if I have a demo, my answer is: "Do you have a smartphone?" Because since mid-2015, Google's speech recognition has been based on LSTM trained by our "connectionist temporal classification". This dramatically improved Google Voice not only by up to ten per cent, but by almost 50 per cent - now available to billions of smartphone users.

AI isn't coming for our jobs, it's coming for our planet and will one day 'colonise the galaxy'

WIRED   R&D

Microsoft's recent ImageNet 2015 winner also uses LSTM-related ideas. The Chinese search giant Baidu is building on our methods, such as CTC. Apple explained at its recent WWDC 2016 developer conference how it is using LSTM to improve iOS. Google is applying the rather universal LSTM not only to speech recognition but also to natural language-processing, machine translation, image caption generation and other fields. Eventually it will end up as one huge LSTM.

AlphaGo, the program that beat the best human Go player, was made by DeepMind, which is influenced by our former students: two of DeepMind's first four members came from my lab.

ADVERTISEMENT

chers. This explains the interest in unsupervised learning (UL). There are two ssive UL is simply about detecting regularities in observation streams. This h fewer computational resources, such as space and time and energy, or data oding, which can be achieved to a certain extent by backpropagation, and can

ian passive UL: it is about learning to shape the observation stream through rning agent figure out how the world works and what can be done in it. Active d creative behaviour in art and music and science and comedy[4], and we have ntists" based on approximations thereof. There is no reason why machines cannot be curious and creative.
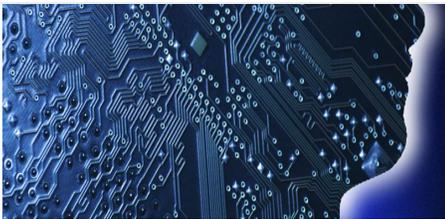
Machine learning versus AI: what's the difference?
—

Machine Learning | 23 Oct 2016

WIRED   R&D

Kids and some animals are still smarter than our best self-learning robots. But I think that within a few year we'll be able to build an NN-based AI (an NNAI) that incrementally learns to become at least as smart as a little animal, curiously and creatively learning to plan, reason and decompose a wide variety of problems into quickly solvable sub-problems.

Once animal-level AI has been achieved, the move towards human-level AI may be small: it took billions of years to evolve smart animals, but only a few millions of years on top of that to evolve humans. Technological evolution is much faster than biological evolution, because dead ends are weeded out much more quickly. Once we have animal-level AI, a few years or decades later we may have human-level AI, with truly limitless applications. Every business will change and all of civilisation will change.

tended sequences using the principle of history compression. Neural Computation, 4(2):234-242.

ng Short-Term Memory. Neural Computation, 9(8):1735-1780. Based on TR FKI-207-95, TUM (1995).

ral networks: An overview. Neural Networks, 61, 85-117.

4. Schmidhuber, J. (2010). Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990-2010). IEEE Transactions on Autonomous Mental Development, 2(3): 230-247, 2010.