

Title: Jürgen Schmidhuber

URL: https://en.wikipedia.org/wiki/J%C3%BCrgen_Schmidhuber

PageID: 405484

Categories: Category:1963 births, Category:Academic staff of the Technical University of Munich, Category:Academic staff of the Università della Svizzera italiana, Category:German artificial intelligence researchers, Category:German computer scientists, Category:Living people, Category:Machine learning researchers, Category:Members of the European Academy of Sciences and Arts, Category:Technical University of Munich alumni

Source: Wikipedia (CC BY-SA 4.0).

Jürgen Schmidhuber (born 17 January 1963) [1] is a German computer scientist noted for his work in the field of artificial intelligence , specifically artificial neural networks . He is a scientific director of the Dalle Molle Institute for Artificial Intelligence Research in Switzerland . [2] He is also director of the Artificial Intelligence Initiative and professor of the Computer Science program in the Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) division at the King Abdullah University of Science and Technology (KAUST) in Saudi Arabia . [3] [4]

He is best known for his foundational and highly-cited [5] work on long short-term memory (LSTM), a type of neural network architecture which was the dominant technique for various natural language processing tasks in research and commercial applications in the 2010s. He also introduced principles of dynamic neural networks , meta-learning , generative adversarial networks [6] [7] [8] and linear transformers , [9] [10] [8] all of which are widespread in modern AI.

Career

Schmidhuber completed his undergraduate (1987) and PhD (1991) studies at the Technical University of Munich in Munich , Germany. [1] His PhD advisors were Wilfried Brauer and Klaus Schulten . [11] He taught there from 2004 until 2009. From 2009, [12] until 2021, he was a professor of artificial intelligence at the Università della Svizzera Italiana in Lugano , Switzerland. [1]

He has served as the director of Dalle Molle Institute for Artificial Intelligence Research (IDSIA), a Swiss AI lab, since 1995. [1]

In 2014, Schmidhuber formed a company, Nnaisense, to work on commercial applications of artificial intelligence in fields such as finance, heavy industry and self-driving cars . Sepp Hochreiter , Jaan Tallinn , and Marcus Hutter are advisers to the company. [2] Sales were under US\$11 million in 2016; however, Schmidhuber states that the current emphasis is on research and not revenue. Nnaisense raised its first round of capital funding in January 2017. Schmidhuber's overall goal is to create an all-purpose AI by training a single AI in sequence on a variety of narrow tasks. [13]

Research

In the 1980s, backpropagation did not work well for deep learning with long credit assignment paths in artificial neural networks . To overcome this problem, Schmidhuber (1991) proposed a hierarchy of recurrent neural networks (RNNs) pre-trained one level at a time by self-supervised learning . [14] It uses predictive coding to learn internal representations at multiple self-organizing time scales. This can substantially facilitate downstream deep learning. The RNN hierarchy can be collapsed into a single RNN, by distilling a higher level chunker network into a lower level automatizer network. [14] [15] In 1993, a chunker solved a deep learning task whose depth exceeded 1000. [16]

In 1991, Schmidhuber published adversarial neural networks that contest with each other in the form of a zero-sum game , where one network's gain is the other network's loss. [6] [17] [7] [8] The first network is a generative model that models a probability distribution over output patterns. The second network learns by gradient descent to predict the reactions of the environment to these

patterns. This was called "artificial curiosity." In 2014, this principle was used in a generative adversarial network where the environmental reaction is 1 or 0 depending on whether the first network's output is in a given set. GANs were the state of the art in generative modeling during 2015-2020 period.

Schmidhuber supervised the 1991 diploma thesis of his student Sepp Hochreiter [18] which he considered "one of the most important documents in the history of machine learning". [15] It studied the neural history compressor , [14] and more importantly analyzed and overcame the vanishing gradient problem . This led to the long short-term memory (LSTM), a type of recurrent neural network . The name LSTM was introduced in a tech report (1995)

leading to the most cited LSTM publication (1997), co-authored by Hochreiter and Schmidhuber. [19] It was not yet the standard LSTM architecture which is used in almost all current applications. The standard LSTM architecture was introduced in 2000 by Felix Gers , Schmidhuber, and Fred Cummins. [20] Today's "vanilla LSTM" using backpropagation through time was published with his student Alex Graves in 2005, [21] [22] and its connectionist temporal classification (CTC) training algorithm [23] in 2006. CTC was applied to end-to-end speech recognition with LSTM. By the 2010s, the LSTM became the dominant technique for a variety of natural language processing tasks including speech recognition and machine translation , and was widely implemented in commercial technologies such as Google Neural Machine Translation , [24] have also been used in Google Voice for transcription [25] and search, [26] and Siri . [27]

In 2014, the state of the art was training "very deep neural network" with 20 to 30 layers. [28] Stacking too many layers led to a steep reduction in training accuracy, [29] known as the "degradation" problem. [30] In May 2015, Rupesh Kumar Srivastava, Klaus Greff, and Schmidhuber used LSTM principles to create the highway network , a feedforward neural network with hundreds of layers, much deeper than previous networks. [8] [31] [32] In Dec 2015, the residual neural network (ResNet) was published, which is a variant of the highway network. [30] [33]

In 1992, Schmidhuber published fast weights programmer , an alternative to recurrent neural networks . [9] It has a slow feedforward neural network that learns by gradient descent to control the fast weights of another neural network through outer products of self-generated activation patterns, and the fast weights network itself operates over inputs. [10] This was later shown to be equivalent to the unnormalized linear Transformer . [34] [10] [35] Schmidhuber used the terminology "learning internal spotlights of attention" in 1993. [36]

In 2011, Schmidhuber's team at IDSIA with his postdoc Dan Ciresan also achieved dramatic speedups of convolutional neural networks (CNNs) on fast parallel computers called GPUs . An earlier CNN on GPU by Chellapilla et al. (2006) was 4 times faster than an equivalent implementation on CPU. [37] The deep CNN of Dan Ciresan et al. (2011) at IDSIA was already 60 times faster [38] and achieved the first superhuman performance in a computer vision contest in August 2011. [39] Between 15 May 2011 and 10 September 2012, these CNNs won four more image competitions [40] [41] and improved the state of the art on multiple image benchmarks. [42] The approach has become central to the field of computer vision . [41] It is based on CNN designs introduced much earlier by Kunihiko Fukushima . [43] [41]

Credit disputes

Schmidhuber has controversially argued that he and other researchers have been denied adequate recognition for their contribution to the field of deep learning , in favour of Geoffrey Hinton , Yoshua Bengio and Yann LeCun , who shared the 2018 Turing Award for their work in deep learning. [2] [44] [45] He wrote a "scathing" 2015 article arguing that Hinton, Bengio and LeCun "heavily cite each other" but "fail to credit the pioneers of the field". [45] In a statement to the New York Times , Yann LeCun wrote that "Jürgen is manically obsessed with recognition and keeps claiming credit he doesn't deserve for many, many things... It causes him to systematically stand up at the end of every talk and claim credit for what was just presented, generally not in a justified manner." [2] Schmidhuber replied that LeCun did this "without any justification, without providing a single example," [46] and published details of numerous priority disputes with Hinton, Bengio and LeCun. [47] [48]

The term "schmidhubered" has been jokingly used in the AI community to describe Schmidhuber's habit of publicly challenging the originality of other researchers' work, a practice seen by some in the AI community as a "rite of passage" for young researchers. Some suggest that Schmidhuber's significant accomplishments have been underappreciated due to his confrontational personality. [49] [44]

Recognition

Schmidhuber received the Helmholtz Award of the International Neural Network Society in 2013, [50] and the Neural Networks Pioneer Award of the IEEE Computational Intelligence Society in 2016 [51] for "pioneering contributions to deep learning and neural networks." [1] He is a member of the European Academy of Sciences and Arts . [52] [12]

He has been referred to as the "father of modern AI" or similar, [62] the "father of Generative AI," [63] and also the "father of deep learning." [64] [57] Schmidhuber himself, however, has called Alexey Grigorevich Ivakhnenko the "father of deep learning," [65] [66] and gives credit to many even earlier AI pioneers. [15]

Views

Schmidhuber is a proponent of open source AI, and believes that they will become competitive against commercial closed-source AI. [8]

Since the 1970s, Schmidhuber wanted to create "intelligent machines that could learn and improve on their own and become smarter than him within his lifetime." [8] He differentiates between two types of AIs: tool AI, such as those for improving healthcare , and autonomous AIs that set their own goals, perform their own research, and explore the universe. He has worked on both types for decades, [8] He expects the next stage of evolution to be self-improving AIs that will succeed human civilization as the next stage in the universal increase towards ever-increasing complexity, and he expects AI to colonize the visible universe. [8]

References

ISNI

VIAF

GND

WorldCat

United States

Israel

Mathematics Genealogy Project

Association for Computing Machinery

zbMATH

Google Scholar

DBLP

DDB

IdRef

Yale LUX