

LSTM 之父，被图灵奖遗忘的大神

人工智能4月2日

大神总是孤独的。

编者按：本文来自微信公众号“新智元”（ID：AI_era），作者张乾、金磊、大明，36 氪经授权发布。

导读：图灵奖为何不颁给 LSTM 之父 Jürgen Schmidhuber？作为 AI 界特立独行的人，Schmidhuber 与深度学习三巨头有过口水战，并现场对质 GAN 的提出者，可谓得罪了一圈人。

一个被遗忘的大神。

前几天，[2018 图灵奖获得者公布](#)，深度学习三巨头：Yoshua Bengio、Geoffrey Hinton、Yann LeCun 荣获桂冠。



HLB (Hinton、LeCun、Bengio) 三人获图灵奖实至名归，消息一出，计算机界纷纷送上祝福。

不过，在恭贺之余，也有不少的网友**发出了质疑**：图灵奖为什么没颁给 LSTM 之父 Jürgen Schmidhuber？他也是深度学习领域的大家啊。



Peter Seeberg @seeberg_peter · 3月28日

Why maybe Geoffrey Hinton, Yoshua Bengio, and Yann LeCun and not **Jürgen Schmidhuber** received the Turing Award.



获得图灵奖的为什么是 Yoshua Bengio、Geoffrey Hinton、Yann LeCun，而不是 Jürgen Schmidhuber 呢？



Abhinav Bhardwaj @codealchemistab · 3月27日

Yann LeCun, Geoffrey Hinton and Yoshua Bengio won the Turing Award which is also known as the "Nobel Prize for Computing" for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

Cool, but what about **Juergen Schmidhuber**?



Yoshua Bengio、Geoffrey Hinton 和 Yann LeCun 获得了图灵奖(计算机界的诺贝尔奖)，让神经网络成为计算机中的重要环节。非常酷！但是 Jürgen Schmidhuber 呢？

就连南京大学人工智能学院院长**周志华教授**也发微博称，LSTM 是教科书级的贡献。



南大周志华

3-27 21:01 来自HUAWEI Mate 20 Pro

+关注

要论对深度学习的贡献，Hinton无疑巨首，LeCun和Schmidhuber贡献都很大。但HLB总捆绑在一起，而S跟HLB都不对劲。。。获奖需有提名有投票，人缘也重要。。。不过没关系，有LSTM这样教科书级的贡献足以淡定//

@Wilson_NJUer:连接主义终于实至名归！不过Jürgen Schmidhuber好像有点小亏 😊 人工智能

Jürgen Schmidhuber 是瑞士 Dalle Molle 人工智能研究所的联合主任，他 1997 年提出的 LSTM 现在被广泛应用在谷歌翻译、苹果 Siri、亚马逊 Alex 等应用中，可谓是深度学习领域最商业化的技术之一。



人工智能

Jürgen Schmidhuber

除了 LSTM 之外，Jürgen Schmidhuber “引以为傲”的还有他在 1992 年提出的一种 PM (Predictability Minimization) 模型。

为什么 “引以为傲要” 打引号？

因为他坚持认为现在大火的 GAN 就是 PM 的变种，两者的区别就在于方向是反的，为此，Jürgen Schmidhuber 还和 GAN 的提出者 Ian Goodfellow 有过线上线下的激烈的交锋，业界至今记忆犹新。

至于对深度学习三巨头 HLB，Jürgen Schmidhuber 也打过几轮口水仗，认为 HLB 三人在自己的圈子里玩，对深度学习领域其他更早期先驱人物的贡献则只字不提，之后 LeCun 发文反击。

Jürgen Schmidhuber 这个暴脾气没得奖，是因为得罪人了吗？

LSTM 之父交战 GAN 之父

故事还得从五年前讲起。

2014 年，Ian Goodfellow 第一篇 GAN 论文投到了 NIPS 大会，三位评审中，两位直接通过，一位直接毙掉。

这位拒稿的评审就是 Jürgen Schmidhuber。

Jürgen Schmidhuber 为什么给出 Goodfellow 这位年轻的后辈如此截然相反的评审意见？

原来，Jürgen Schmidhuber 认为，GAN 不能称为第一个对抗网络，他自己在 1992 年提出的 PM 模型才是。

LEARNING FACTORIAL CODES BY PREDICTABILITY MINIMIZATION

(*Neural Computation*, 4(6):863–879, 1992)

Jürgen Schmidhuber

Department of Computer Science
University of Colorado
Campus Box 430, Boulder, CO 80309, USA
yirgan@cs.colorado.edu

Abstract

I propose a novel general principle for unsupervised learning of distributed non-redundant internal representations of input patterns. The principle is based on two opposing forces. For each representational unit there is an adaptive predictor which tries to predict the unit from the remaining units. In turn, each unit tries to react to the environment such that it minimizes its predictability. This encourages each unit to filter 'abstract concepts' out of the environmental input such that these concepts are statistically independent of those upon which the other units focus. I discuss various potentially powerful implementations of the principle which aim at finding binary factorial codes (Bar-

在给 Goodfellow 的评审意见中，Jürgen Schmidhuber 直接质疑 Goodfellow :GAN 和 PM 在许多方面看起来很相似。这两种方法都使用“对抗性”MLP 来估计某些概率，并学习编码分布。不同之处在于，新系统学会根据统计独立的随机输入生成非平凡分布，而旧的 PM 学习生成统计独立的随机输出以响应非平凡分布(通过提取相互独立的因子特征编码分布)。

因此，GAN 本质上改变了 PM 的方向——这是主要的区别吗？GAN 应该被称为“反向 PM”吗？

In the context of backpropagation "into a stochastic network with continuous activations (so that gradients can be back-propagated)," one should also cite the much older original papers on this, in particular those by Williams (e.g., Machine Learning, 1992), who adjusted mean and variance of probabilistic continuous units by backpropagation through random number generators.

Finally, how is the submission related to the first work on "adversarial" MLPs for modeling data distributions through estimating conditional probabilities, which was called "predictability minimisation" or PM (Schmidhuber, NECO 1992)? The new approach seems similar in many ways. Both approaches use "adversarial" MLPs to estimate certain probabilities and to learn to encode distributions. A difference is that the new system learns to generate a non-trivial distribution in response to statistically independent, random inputs, while good old PM learns to generate statistically independent, random outputs in response to a distribution by extracting mutually independent, factorial features encoding the distribution). Hence the new system essentially inverts the direction of PM. Should it perhaps be called "inverse PM"?

最后，无奈之下的 Goodfellow 只好在论文的最终版本里加入了 GAN 和 PM 不同之处的比较，这才有了第一篇 GAN 论文的诞生。

然而 Jürgen Schmidhuber 依旧不依不饶，私下里通过邮件跟 Goodfellow 几番争论。

最激烈的事情发生在 2016 年。

当时 GAN 已经在学术界名气日旺，2016 年 NIPS 大会，Goodfellow 在现场有个 Tutorial，正当他讲到 GAN 与其他模型相比较时，被一个听众的提问打断了。

这个听众就是 Jürgen Schmidhuber。



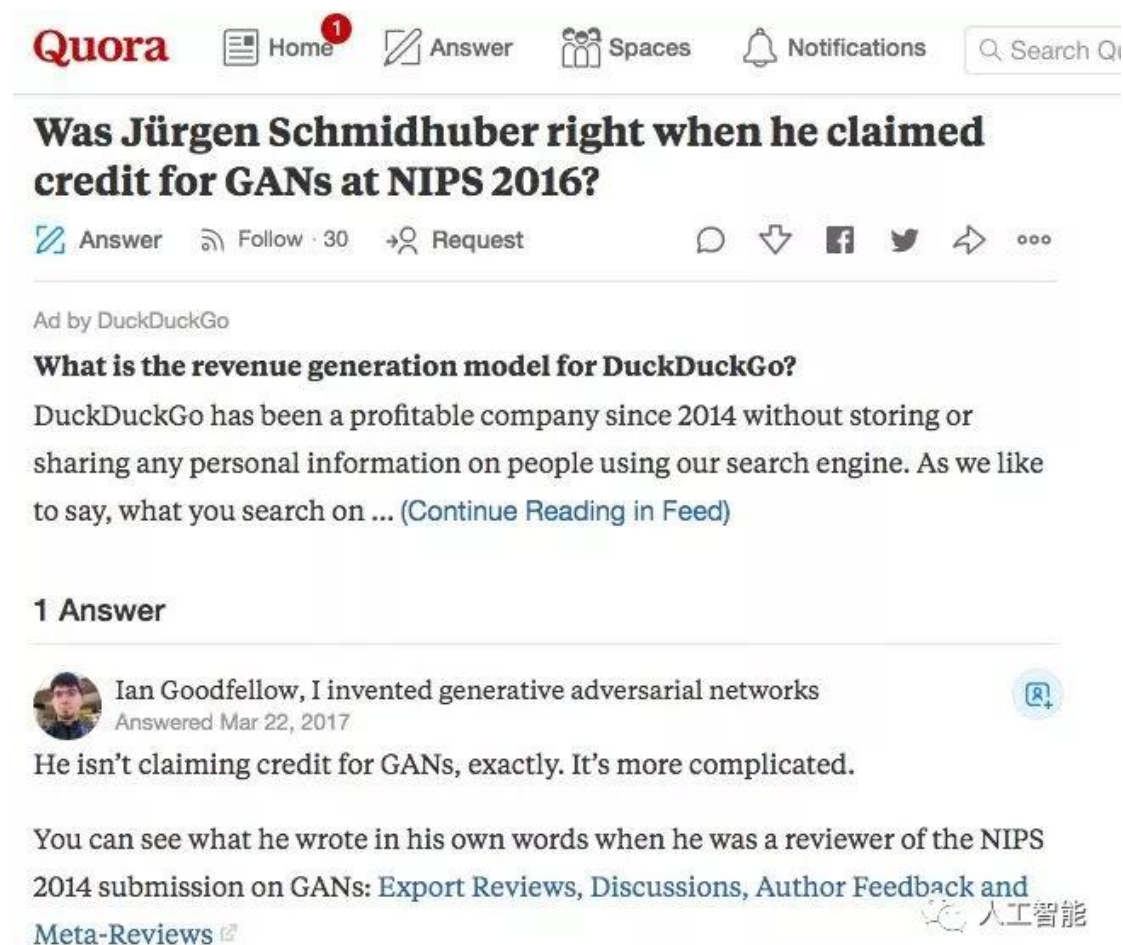
Jürgen Schmidhuber 向 Goodfellow 提问

Jürgen Schmidhuber 的问题很长，大概说了两分钟，主要内容是强调说自己在 1992 年就已经提出来 PM，接着说了一大堆它的原理、实现过程等等，最后图穷匕见：你说说你的 GAN 和我的 PM 有没有相似之处？

Goodfellow 也不示弱：你说的问题我们之前在邮件里已经交流过很多次了，我也早就公开回应过你了，不想在现在的场合浪费听众的耐心。（掌声）

一位五十多岁的长者试图碾压三十出头的小伙子，但被小伙子反杀，场面一度十分尴尬。

后来，Goodfellow 在 Quora 上透露，他联系了 NIPS 的主办方，询问 Jürgen Schmidhuber 是否有办法向他提出投诉，并由 NIPS 代表委员会判断 Goodfellow 的论文是否不公平。但主办方说并没有这样的流程。



The screenshot shows a Quora page with the following elements:

- Navigation Bar:** Quora logo, Home (with a red notification badge), Answer, Spaces, Notifications, and a search bar.
- Question:** "Was Jürgen Schmidhuber right when he claimed credit for GANs at NIPS 2016?"
- Actions:** Answer, Follow (30), Request, and social sharing icons (comment, download, Facebook, Twitter, share, and more).
- Ad:** "Ad by DuckDuckGo" with the title "What is the revenue generation model for DuckDuckGo?". The ad text states: "DuckDuckGo has been a profitable company since 2014 without storing or sharing any personal information on people using our search engine. As we like to say, what you search on ... (Continue Reading in Feed)".
- Answer:** "1 Answer" section with one answer by Ian Goodfellow. The answer text is: "He isn't claiming credit for GANs, exactly. It's more complicated. You can see what he wrote in his own words when he was a reviewer of the NIPS 2014 submission on GANs: [Expert Reviews, Discussions, Author Feedback and Meta-Reviews](#)".

除此之外，Jürgen Schmidhuber 还提出和 Goodfellow 一起写合一篇描述 PM 和 GAN 之间相同点和不同点的论文，但前提是两人真正能达成一致。现在看起来，这是不太可能了。

交恶深度学习三巨头

Ian Goodfellow 是 Bengio 的得意门生，而在惹 Goodfellow 之前，Jürgen Schmidhuber 已经与 Goodfellow 的师傅 Bengio 以及 HLB 三人组合有过口水战。

2015年5月，Bengio、Hinton 和 LeCun 三位大神联手在 Nature 上发了一篇 review，题目直接就叫《Deep Learning》。此文从传统的机器学习技术讲起，总结了现代机器学习的主要架构和方法，描述了训练多层网络架构的反向传播算法，以及卷积神经网络的诞生，分布式表示和语言处理，以及递归神经网络及其应用等等。

REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun^{1,2}, Yoshua Bengio³ & Geoffrey Hinton^{4,5}

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

人工智能

这篇文章堪称是深度学习领域的经典之作，Google 学术统计的统计显示，论文被引用数已接近 14000 次。

从这篇文章的行文风格和内容来看，相当一部分内容是三位作者赖以成名的划时代成就，大有为深度学习“回顾历史、展望未来”的意思，字里行间透露出“盖棺定论”的豪迈之意挡也挡不住。

不过 Nature 上这篇文章发表后不到一个月，他就在自己的博客上发文，对这篇论文进行了批评。



Critique of Paper by "Deep Learning Conspiracy" (Nature 521 p 436)

Jürgen Schmidhuber

Pronounce: You_again Shmidhoobuh
June 2015

人工智能

Schmidhuber 在文中表示，这篇文章让他非常不爽，因为全文多次引用三位作者自己的研究成果，而对于其他先驱人物对深度学习更早的贡献则只字不提，比如：

- 三位作者俨然以 AI 先驱自居，在参考文献信息中**对深度学习之父 Alexey Grigorevich Ivakhnenko 根本提都没提**，后者早在 1965 年就发表了第一篇面向深度网络的通用深度学习算法的论文。1971 年的论文中就提出了 8 层深度神经网络。
-
- 文章提到了反向传播 (BP)，但**引用的都是自己的论文**，根本没有提到反向传播的发明者和早期开拓者的成果。实际上，最早的反向传播模型诞生于上世纪 60-70 年代。

-
-

关于前馈神经网络，Review 说是 CIFAR 实验室的研究人员 2006 年努力导致 FNN 的再度复兴，这里又是在自夸，而且是一种误导。实际上，研究人员已经使用 Ivakhnenko 几十年了。

-
-

文章在提到无监督预训练 FNN 时还是在引用作者自己的研究，但却没提 **Schmidhuber 本人早在 1992-1993 年就提出的无监督式预训练 RNN**，只不过那时候还不叫 RNN，但原理和思想是一致的。

-
-

在说到非监督学习对深度学习复兴的深远影响时，仍然只引用了作者们自己的成果。

-
-

同样，在卷积神经网络一节，文章提到了“池化”，但却没有提及提出最大池化技术的先驱人物等等。

-

总之，在 Schmidhuber 这篇文章中认为，今年获得图灵奖的“深度学习三巨头”俨然成了贪他人之功，以为己利的鸡贼、借助江湖地位互相吹捧，压制老前辈的学阀。这一篇文章的操作也真是刚的很。

并且，Schmidhuber 还发表过 Hinton、LeCun 等人出名，是因为背后有谷歌、Facebook 这样的大公司在背后做宣传的言论。

后来，LeCun 在一封 email 回复中写道：“Jürgen 对众人的认可过于痴迷，**总是说自己没有得到应得的很多东西**。几乎是惯性地，他总是在别人每次讲话结束时都要站起来，说刚刚提出的成果有他的功劳，大体上看，这种行为并不合理。”

由此看来，Schmidhuber 和三巨头这梁子算是结下了。

被图灵奖遗忘的大神？

1997 年，Jürgen Schmidhuber 和 Sepp Hochreiter 发表了一篇关于一种**递归神经网络**的论文，就是大名鼎鼎的长短期记忆网络（LSTM）。

LONG SHORT-TERM MEMORY

NEURAL COMPUTATION 9(8):1735–1780, 1997

Sepp Hochreiter
Fakultät für Informatik
Technische Universität München
80290 München, Germany
hochreit@informatik.tu-muenchen.de
<http://www7.informatik.tu-muenchen.de/~hochreit>

Jürgen Schmidhuber
IDSIA
Corso Elvezia 36
6900 Lugano, Switzerland
juergen@idsia.ch
<http://www.idsia.ch/~juergen> 人工智能

2015 年，LSTM 被用于谷歌智能手机软件中的语音识别新实现。谷歌也使用 LSTM 作为智能助手 Allo 和谷歌翻译。后来苹果在 iPhone 和 Siri 的“Quicktype”功能中使用了 LSTM。Amazon 的 Alexa 也使用了 LSTM。在 2017 年，Facebook 每天使用 LSTM 网络进行约 45 亿次自动翻译，LSTM 可以说是商业化做广泛的 AI 技术之一。

除了 LSTM，2011 年 Jürgen Schmidhuber 还与他的博士后学生在 GPU 上实现 CNN(卷积神经网络)的显著加速，现在这种方法已经成为**计算机视觉领域的核心**。

当今年的图灵奖公布之后，有不少人认为单纯从贡献上来讲，提出 LSTM 的 Jürgen Schmidhuber 也应该获奖。



Kostas Kamnitsas @KostasKamnitsas · 3月27日

Congratulations to all three, very well deserved. But the massive contributions of **Jürgen Schmidhuber** being left out is a big surprise and wrong.

人工智能

祝贺三位，当之无愧。但贡献如此显著的 Jürgen Schmidhuber 被排除在外的还是非常意外和错误的。

京东集团副总裁、加拿大 Simon Fraser 大学计算科学学院教授、统计与精算系教授、加拿大一级研究讲座教授、ACM Fellow、IEEE Fellow、ACM SIGKDD Chair 裴健教授告诉新智元：图灵奖揭晓后，经常有人有疑问，那个谁谁谁对这个方向这个领域也作出了突出贡献，甚至比其中某一位获奖者贡献还大，为什么没有一起获奖？然后就顺理成章的地有各种揣测。

“我个人的理解是，每一届图灵奖是授予一个个人或一个团队，而不是给一个领域相对独立的多位先驱。“The ACM A. M. Turing Award is an annual prize given by the Association for Computing Machinery (ACM) to an individual selected for contributions ‘of lasting and major technical importance to the computer field’ .”

图灵奖历史上多次发生过在同一领域独立作出突出贡献的多位科学家和团队陆续获奖的佳话，如计算复杂性理论和数据库理论等都有多位获奖者。

所以，各位看官不要着急，也不要以看宫廷戏的心态看图灵奖。

OpenCV 的创造者、AI 科学家 Gary Bradski 曾评价 Schmidhuber：“他是做了很多开创性的工作，但他不是那个使这些成果流行起来的人。这就像最早发现美洲的是维京人，但千古留名的是哥伦布。”

大神总是孤独的

Jürgen Schmidhuber 和 HLB 各有各的成就，但大神们总有相似之处，那就是坚持。

Hinton 在上大学时期坚信神经网络是未来，并为此坚持了三十年。

出生于 1963 年的 Jürgen Schmidhuber，在他 15 岁时就坚信通用人工智能一定会实现，当时那个年代中国刚刚改革开放。



婴儿时期的 Jürgen Schmidhuber，左边是他父亲

“作为一个少年，我意识到人们可以做的最重要的事情就是建立一种学会变得比人类聪明的东西。”

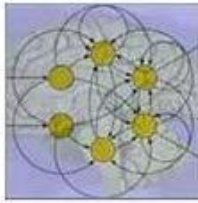
Jürgen Schmidhuber 在年轻时候曾对他弟弟讲，人类可以一个原子一个原子地重建大脑，可以用铜线代替我们缓慢的神经元作为连接，想象力非常大胆。弟弟刚开始很

反对哥哥这套人造大脑可以模仿人类情感和自由意志的观点。但最终，“我意识到他是正确的。”

高中毕业后，Schmidhuber 从 1981 年开始学习计算机科学和数学，并在西德军队服役 15 个月，服兵役时就展现出他特立独行的性格，他不喜欢被人颐指气使，特别是被要求做一些他认为没用的事情。

他的在线简历精心编制了他在学术界的历程，包括诸如“加州理工学院拒绝了他的博士后申请”等等，到现在，Schmidhuber 的个人主页也是颇有特色。

RESEARCH



Artificial Recurrent Neural Networks (1989-2014). Most work in machine learning focuses on machines with reactive behavior. RNNs, however, are more general sequence processors

inspired by human brains. They have adaptive feedback connections and are in principle as powerful as any computer. The first RNNs could not learn to look far back into the past. But our "Long Short-Term Memory" (LSTM) RNN overcomes this fundamental problem, and efficiently learns to solve many previously



unlearnable tasks. It can be used for speech recognition, time series prediction, music composition, etc. In 2009, our LSTM RNNs became the first recurrent Deep Learning systems to win

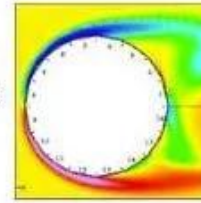
official international competitions (with secret test set known only to the organisers) - they outperformed all other known methods on the difficult problem of recognizing unsegmented cursive handwriting, and also on aspects of



speech recognition. They learn through gradient descent and / or evolution or both. Compare the RNN Book Preface. LSTM is getting popular: Google, Apple, Microsoft, Facebook, IBM, Baidu,

and many other companies use LSTM RNNs to improve large vocabulary speech recognition, machine translation, language identification / time series prediction / text-to-speech synthesis, etc.

Artificial Evolution. State-of-the-art methods for network evolution co-evolve all neurons in parallel (excellent results in various applications). EVOLINO outperforms previous methods on several



supervised learning tasks, and yields the first recurrent support vector machines. Probabilistic incremental program evolution evolves computer programs through probabilistic templates instead of program populations (first approach to evolving entire soccer team strategies from



scratch). As an undergrad Schmidhuber also implemented the first genetic programming system with loops and variable length code (1987, see below). Our novel Natural Evolution Strategies (2008-) yield

excellent results and link policy gradients to evolution. And while most previous algorithms can evolve only hundreds of adaptive parameters, but not millions, our Compressed Network Search (1995-) finds compact descriptions of huge networks. A 2013 variant was the first method to evolve neural network controllers with over a million weights. Compare work on learning to think.

Interestingness & Active Exploration & Artificial Curiosity & Theory of Surprise (1990-2010). Schmidhuber's curious learning agents like to go where they expect to learn something. These rudimentary artificial



Schmidhuber 对通用人工智能的追求一直持续了 40 多年,他还梦想建立一个智能机器劳动的乌托邦,因此在 1988 年,他捐赠了数百万美元来创建瑞士的 Dalle Molle 人工智能研究所。它与当地大学的合作,加上政府源源不断的资助,帮助这座小镇变成了人工智能的天堂中心。

通用 AI 将掌管并改变整个宇宙

《硅谷钢铁侠》的作者在 2018 年 5 月写过一篇 Jürgen Schmidhuber 的特稿，题目是《这个人是 AI 圈想要忘记的教父》。

Bloomberg Businessweek

This Man Is the Godfather the AI Community Wants to Forget

Jürgen Schmidhuber says he'll make machines smarter than us.
His peers wish he'd just shut up.

By Ashlee Vance

 人工智能

这篇文章提到，在大多数学术界之外，Schmidhuber 仍然很不为人知。主要是因为学术圈里的同伴不喜欢他，不少同行评价他自私、狡猾，给人带来痛苦。

由于 Schmidhuber 频繁在学术期刊和会议上怼研究人员，打断他人演讲要求同行承认他们借用甚至窃取了他的想法，后来业内创造了一个动词“Schmidhubered”，谁被别人攻击了就可以用 Schmidhubered。

Schmidhuber 被边缘化的另一部分原因是因为他的研究所位于阿尔卑斯山，位置孤立，远离科技大公司。

2013 年，Schmidhuber 还在瑞士创办了一家创业公司 Nnaisense，这家公司肩负着实现通用人工智能的使命，并影响到 DeepMind。

TEAM



Jürgen Schmidhuber
Co-founder
Chief Scientist



Faustino Gomez
Co-founder
CEO



Jan Koutník
Co-founder
Director of Intelligent
Automation



Bas Steunebrink
Co-founder
Director of Artificial General
Intelligence



Jonathan Masci
Co-founder
Director of Deep Learning

人工智能

Nnaisense 公司主要成员

DeepMind 联合创始人 Shane Legg 以及首批员工之一 Daan Wierstra 就是 Schmidhuber 的学生，他的一些其他博士也加入了 DeepMind。

Jürgen Schmidhuber 曾说，他从 15 岁起就决定，创造出比人类更聪明的机器人然后退休，并把这个观点一致保持到了现在。

他认为，拥有自我意识的 (self-aware) 或 “有知觉的机器” (conscious machines) 不久就将出现。这个观点更加剧了同行对他的不屑。而对这场辩论需要提出一个问题：人工智能是一个工程学科，还是一场创造新的超智能生物的 “造神运动” ？

Schmidhuber 坚定地站在造神的立场，他认为这些技术的基本概念已经存在，而且人类的意识并不神奇，认为机器的意识将从更强大的计算机和算法中出现，而这些算法与他早已设计好的那些非常接近。

这种信念背后，是他坚定不移的认为：**我们生活在《黑客帝国》形式的(Matrix-style)计算机模拟中。**

“在我的成长的过程中，我一直追问自己，我能产生的最大的影响是什么？”

Schmidhuber 博士回忆道，“这个问题后来变得清楚了，就是**我要造出比我自己更聪明的东西，而这个东西又将造出更聪明的东西，如此等等，最终它将掌管并改变整个宇宙，使整个宇宙变得智能。**”

今天，他已经不再困惑这样的机器是否会出现，他说，只要计算能力得到大飞跃，很快就会出现。

参考资料：

<https://www.bloomberg.com/news/features/2018-05-15/google-amazon-and-facebook-owe-j-rgen-schmidhuber-a-fortune>

<https://www.quora.com/Was-J%C3%BCrgen-Schmidhuber-right-when-he-claimed-credit-for-GANs-at-NIPS-2016>

<https://www.inverse.com/article/25521-juergen-schmidhuber-ai-consciousness>