

# Machine Learning

lanyanyan@ict.ac.cn



# Self-Introduction: Yanyan Lan(兰艳艳)

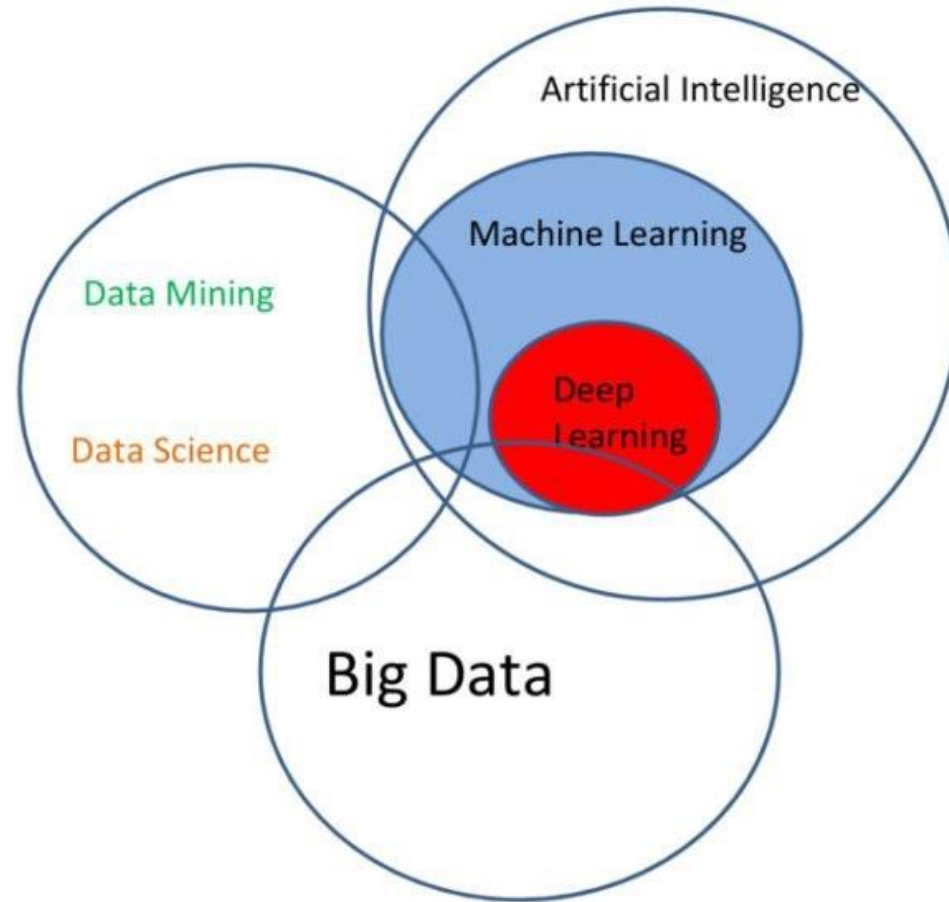
- Associate Professor in Institute of Computing Technology, Chinese Academy of Sciences, a research group leader in (CAS) Key Lab of Network Data Science and Engineering
- Research Interest: Machine Learning, Information Retrieval, Data Mining, and Applications.
  - 2006-2012: Learning to Rank (Search & Recommendation)
  - 2013-2018: Deep Learning for Text Data Analysis (Search, QA and Dialogue)
  - 2018 till now: Learning to Reasoning (QA and Dialogue)
- Homepage: [www.bigdatalab.ac.cn/~lanyanyan](http://www.bigdatalab.ac.cn/~lanyanyan)
- 12 hours' course
  - Statistical Learning Framework, Supervised Learning, SVM

Survey

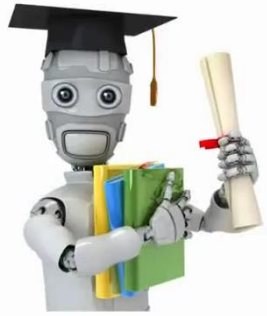
# Some Suggestions

- Teaching
  - Top 30%
  - English Slides
  - Elements
- Requirements
  - Knowledge of **basic computer science principles and skills**
  - Familiarity with the **probability theory**
  - Familiarity with **linear algebra**
- Target
  - Basic: Exam
  - Engineering: Practice
  - Research: Theory + Practice

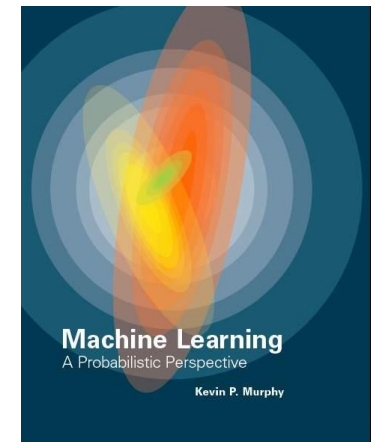
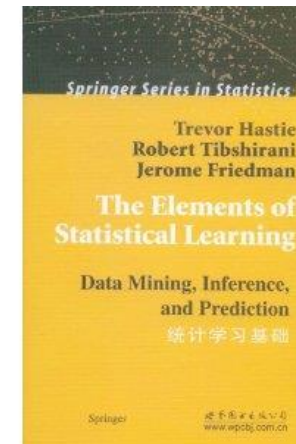
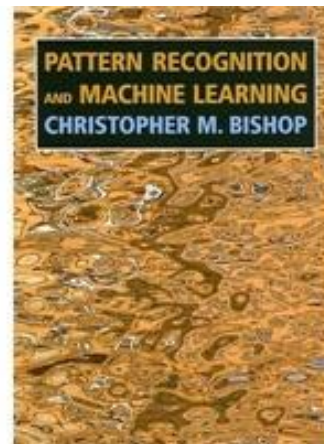
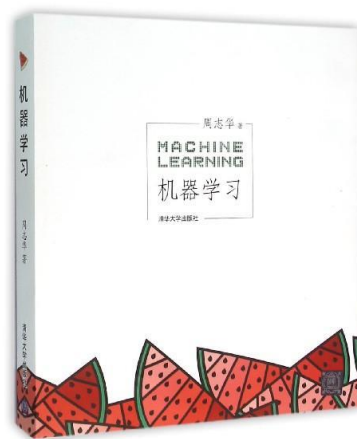
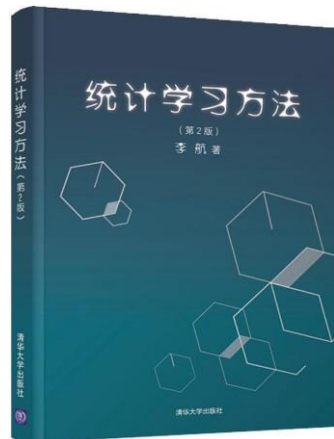
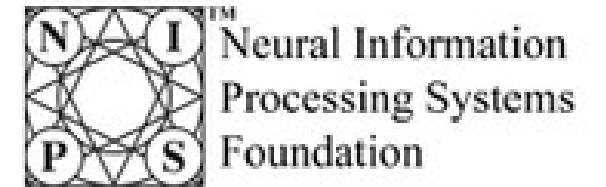
# Relations with Other Fields



# Courses, Books, and Conferences



Machine Learning



# Courses, Books, and Conferences

- **Stanford: CS229**  
<http://cs229.stanford.edu>
- **Berkeley: CS 189/289A**  
<https://people.eecs.berkeley.edu/~jrs/189/>
- **CMU: 10701**  
<http://www.cs.cmu.edu/~epxing/Class/10701/>
- **Cambridge: Information Theory, Pattern Recognition and Neural Networks**  
<http://www.inference.phy.cam.ac.uk/mackay/itprnn/>



# What is Machine Learning?

- Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. **Machine learning explores the study and construction of algorithms that can learn from and make predictions on data.** Such algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions, rather than following strictly static program instructions.
- 1959, Arthur Samuel: "Field of study that gives computers the ability to learn without being explicitly programmed".
- Tom M. Mitchell: "A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ ". (**PAC theory**)



# Defining the Learning Task

Improve on task,  $T$ , with respect to  
performance metric,  $P$ , based on experience,  $E$ .

$T$ : Categorize email messages as spam or legitimate.

$P$ : Percentage of email messages correctly classified.

$E$ : Database of emails, some with human-given labels

$T$ : Recognizing hand-written words

$P$ : Percentage of words correctly classified

$E$ : Database of human-labeled images of handwritten words

$T$ : Driving on four-lane highways using vision sensors

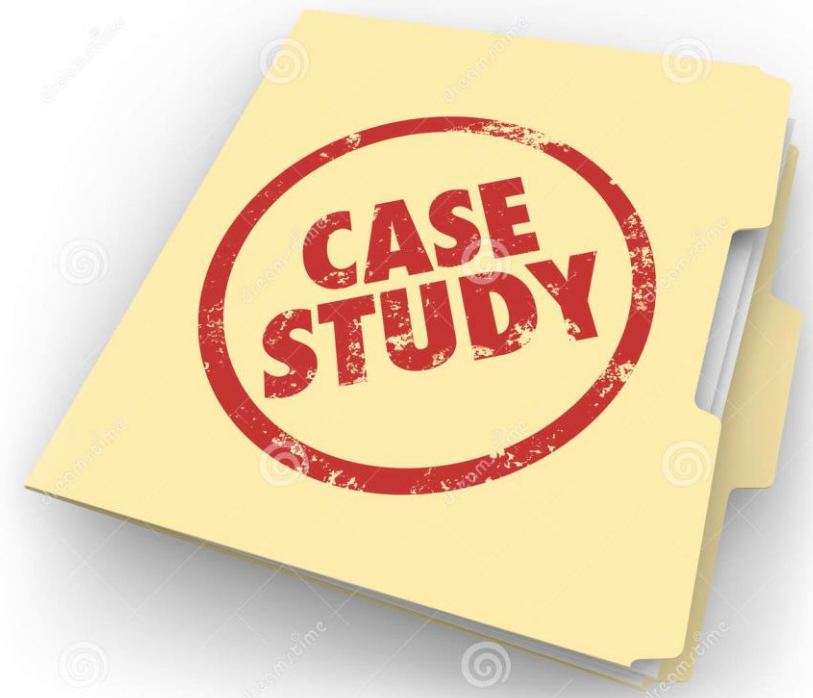
$P$ : Average distance traveled before a human-judged error

$E$ : A sequence of images and steering commands recorded while observing a human driver.

$T$ : Playing checkers

$P$ : Percentage of games won against an arbitrary opponent

$E$ : Playing practice games against itself



# A Learning Comic



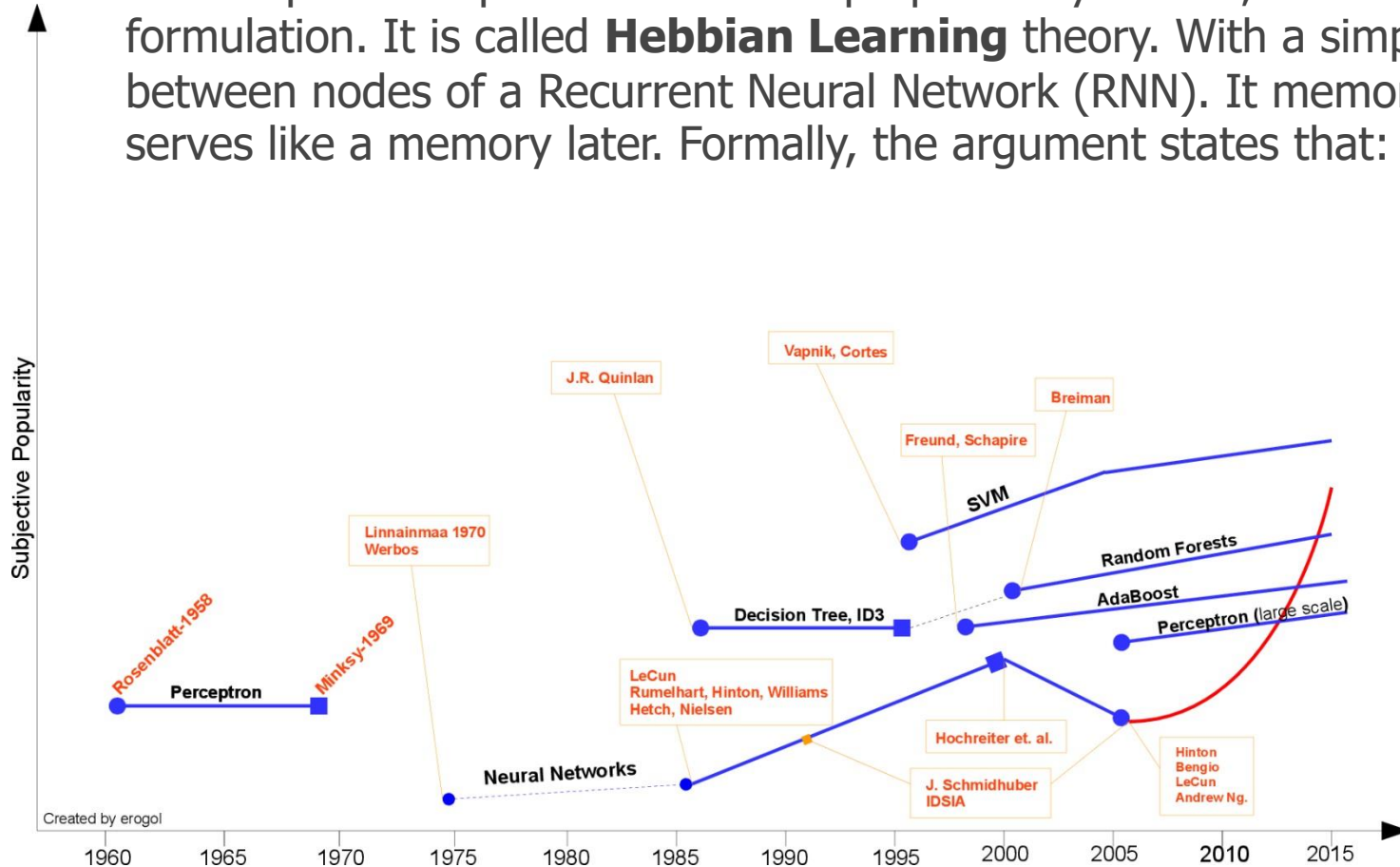
- Experience: Training Data
- What is to be learned: Target Function
- Learning Algorithm: how to infer the target function from the experience
- Evaluation: Test Data

# Categories of Machine Learning

- **Supervised Learning**: direct labeled training data
- **Unsupervised Learning**: unlabeled training data
- **Semi-Supervised Learning**: unlabeled + labeled training data
- **Reinforcement Learning**: indirect labeled training data
- **Transfer Learning**: training and test data are Non IID
- **Multi-Task Learning**: multiple task share representation
- **Active Learning**: actively choose training data

# History of Machine Learning (From Eren Golge's Blog)

First step toward prevalent ML was proposed by **Hebb**, in 1949, based on a neuropsychological learning formulation. It is called **Hebbian Learning** theory. With a simple explanation, it pursues correlations between nodes of a Recurrent Neural Network (RNN). It memorizes any commonalities on the network and serves like a memory later. Formally, the argument states that:

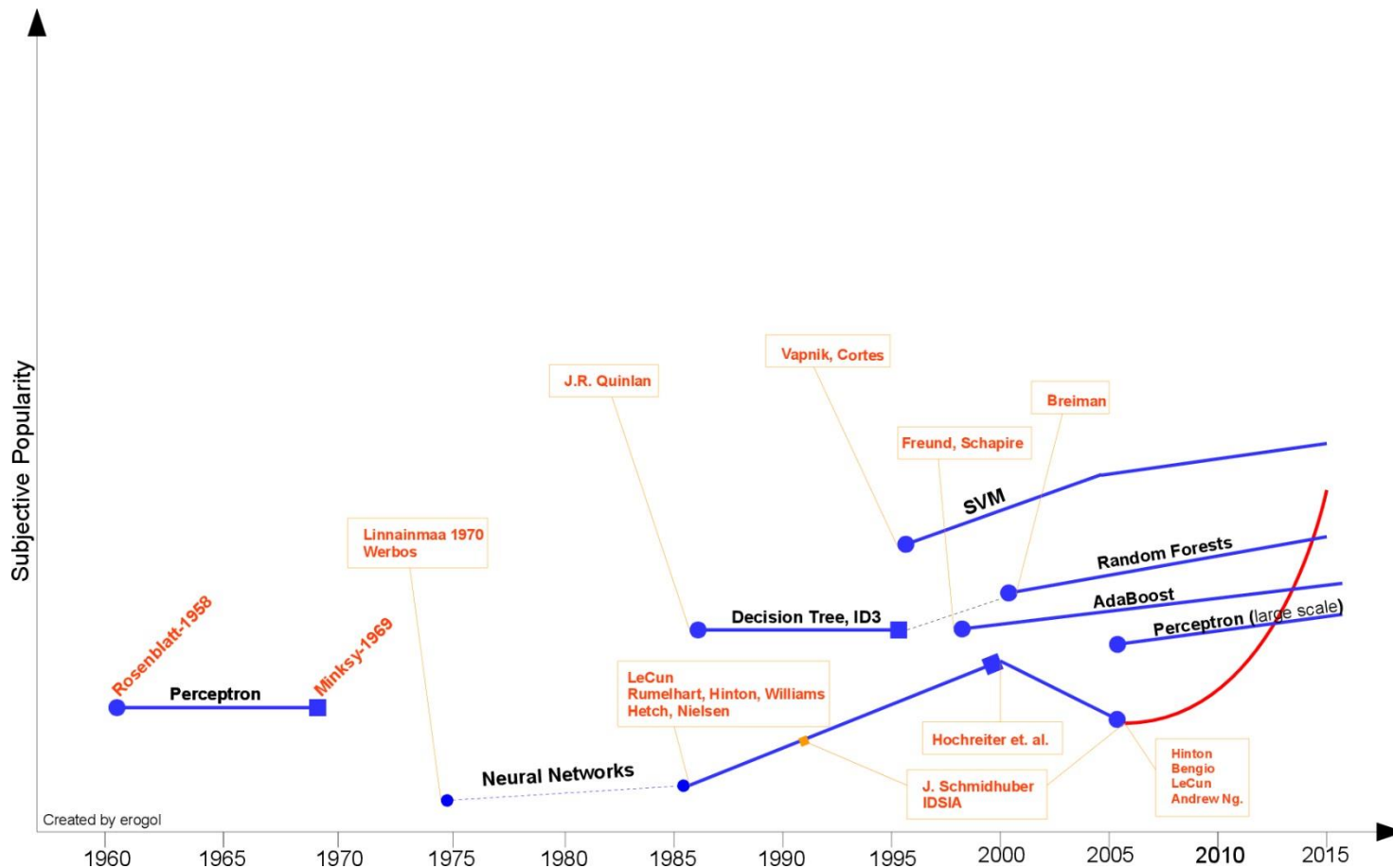


Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability.... When an axon of cell  $A$  is near enough to excite a cell  $B$  and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that  $A$ 's efficiency, as one of the cells firing  $B$ , is increased.

# History of Machine Learning (From Eren Golge's Blog)



Arthur Samuel



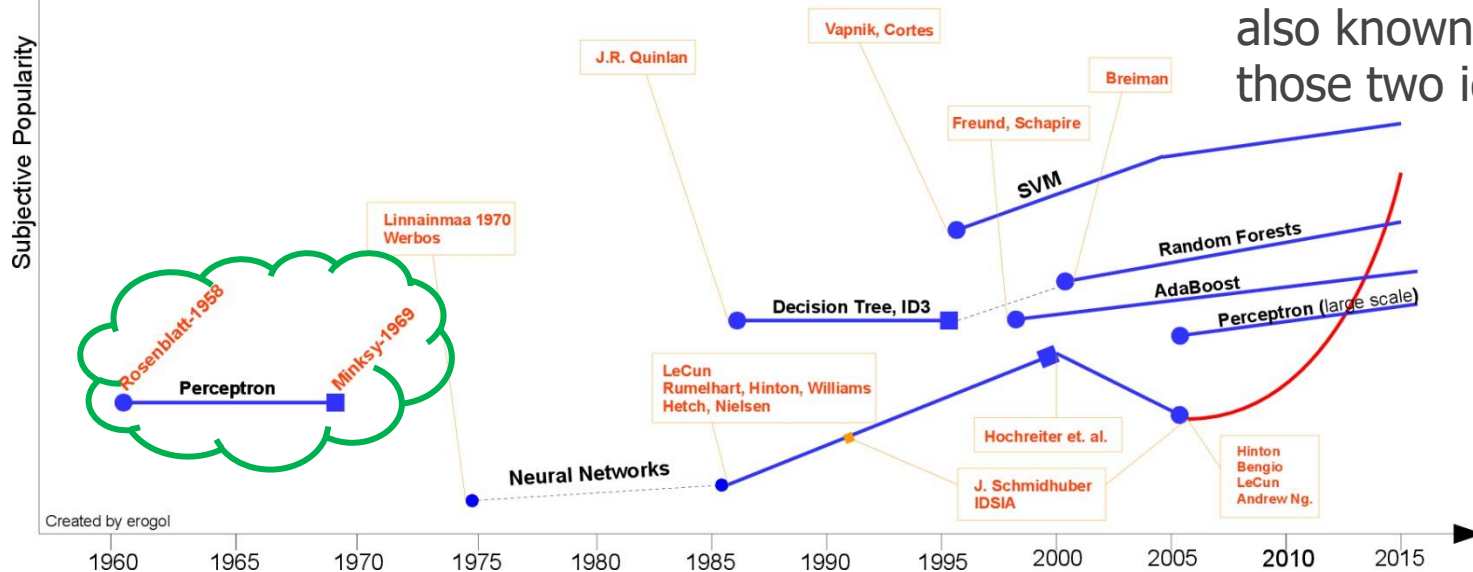
In 1952, Arthur Samuel at IBM, developed a program playing **Checkers**. The program was able to observe positions and learn an implicit model that gives better moves for the latter cases. Samuel played so many games with the program and observed that the program was able to play better in the course of time.



# History of Machine Learning (From Eren Golge's Blog)

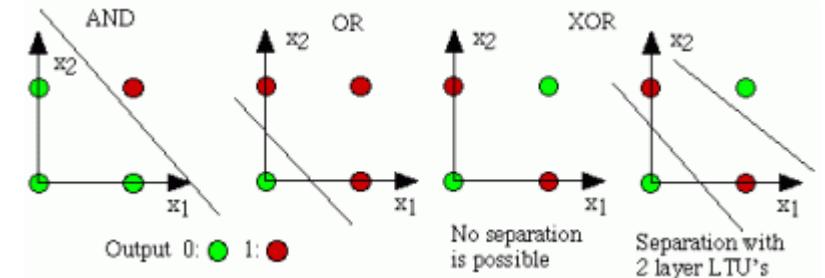
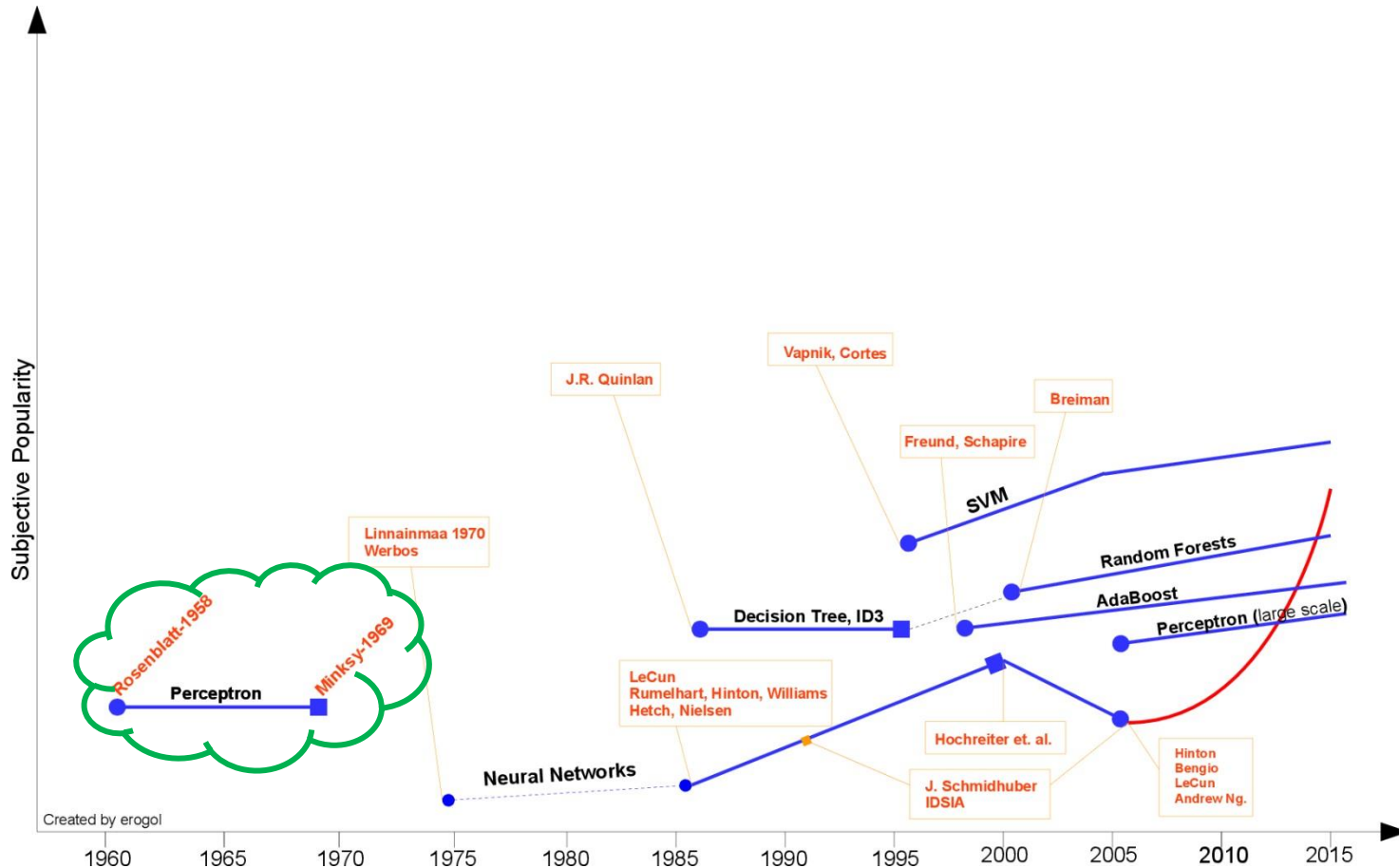
**In 1957 , Rosenblatt's Perceptron** was the second model proposed again with neuroscientific background and it is more similar to today's ML models. It was a very exciting discovery at the time and it was practically more applicable than Hebbian's idea. The perceptron is designed to illustrate some of the fundamental properties of intelligent systems in general, without becoming too deeply enmeshed in the special, and frequently unknown, conditions which hold for particular biological organisms.

**Widrow** engraved **Delta Learning rule** that is then used as practical procedure for Perceptron training. It is also known as **Least Square** problem. Combination of those two ideas creates a good linear classifier.



F. Rosenblatt

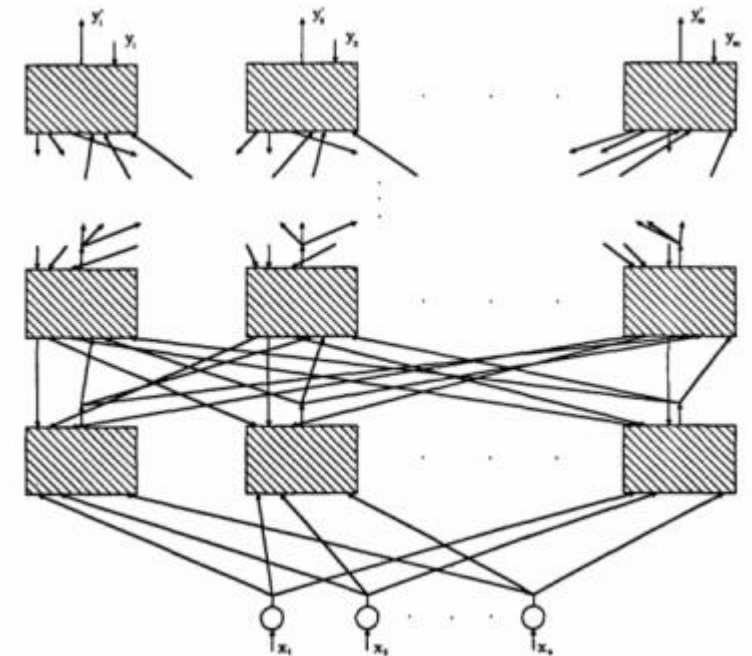
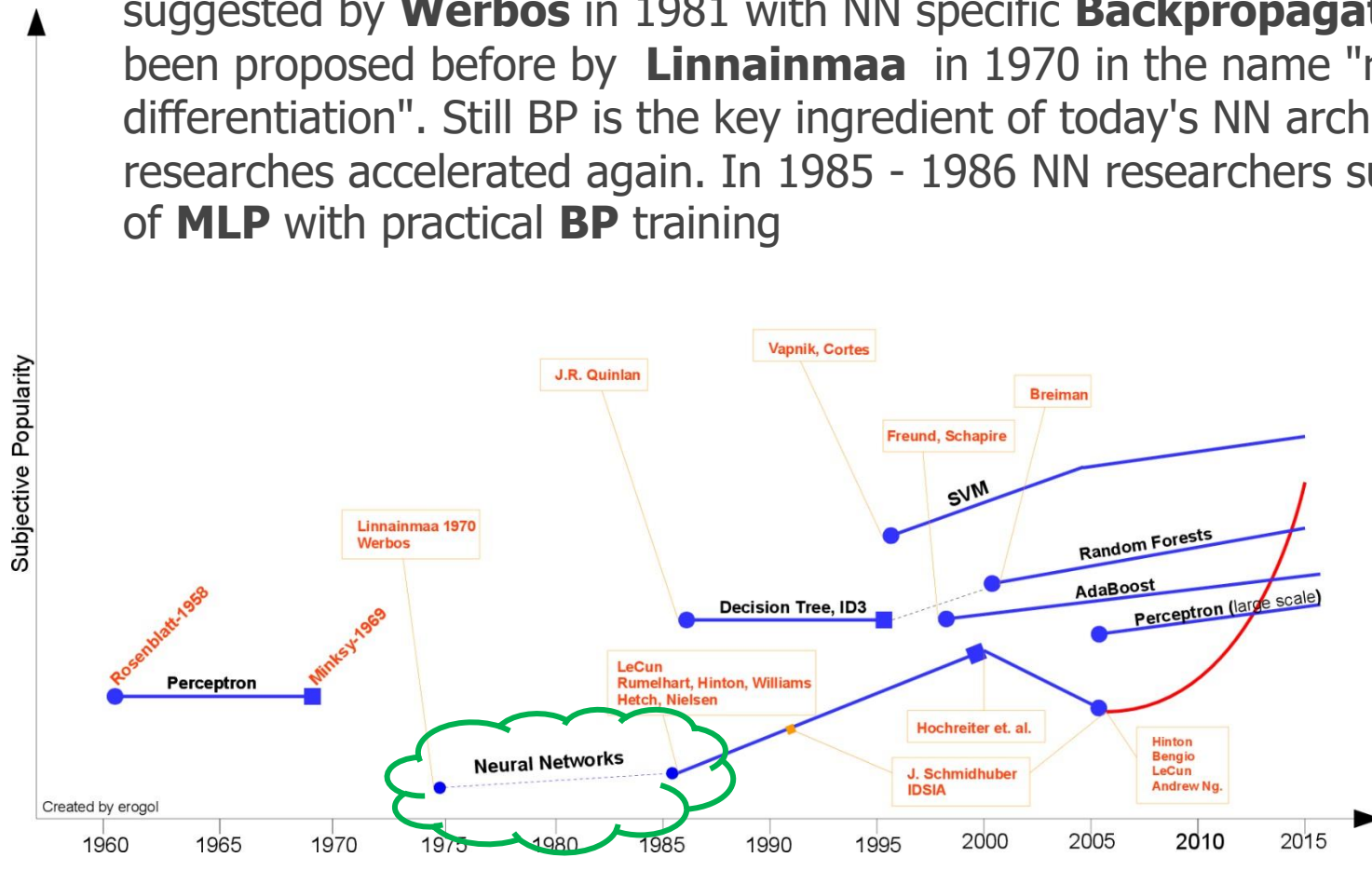
# History of Machine Learning (From Eren Golge's Blog)



However, Perceptron's excitement was hinged by **Minsky** in 1969 . He proposed the famous **XOR** problem and the inability of Perceptrons in such linearly inseparable data distributions. It was the Minsky's tackle to NN community. Thereafter, NN researches would be dormant up until 1980s

# History of Machine Learning (From Eren Golge's Blog)

There had been not to much effort until the intuition of **Multi-Layer Perceptron (MLP)** was suggested by **Werbos** in 1981 with NN specific **Backpropagation(BP)** algorithm, albeit BP idea had been proposed before by **Linnainmaa** in 1970 in the name "reverse mode of automatic differentiation". Still BP is the key ingredient of today's NN architectures. With those new ideas, NN researches accelerated again. In 1985 - 1986 NN researchers successively presented the idea of **MLP** with practical **BP** training



Hecht-Nielsen, Robert. "Theory of the backpropagation neural network." *Neural Networks, 1989. IJCNN, International Joint Conference on*. IEEE, 1989.



# History of Machine Learning (From Eren Golge's Blog)

At the another spectrum, a very-well known ML algorithm was proposed by **J. R. Quinlan** in 1986 that we call **Decision Trees** , more specifically **ID3** algorithm. This was the spark point of the another mainstream ML. Moreover, ID3 was also released as a software able to find more real-life use case with its simplistic rules and its clear inference, contrary to still black-box NN models. After ID3, many different alternatives or improvements have been explored by the community (e.g. ID4, Regression Trees, CART ...) and still it is one of the active topic in ML.

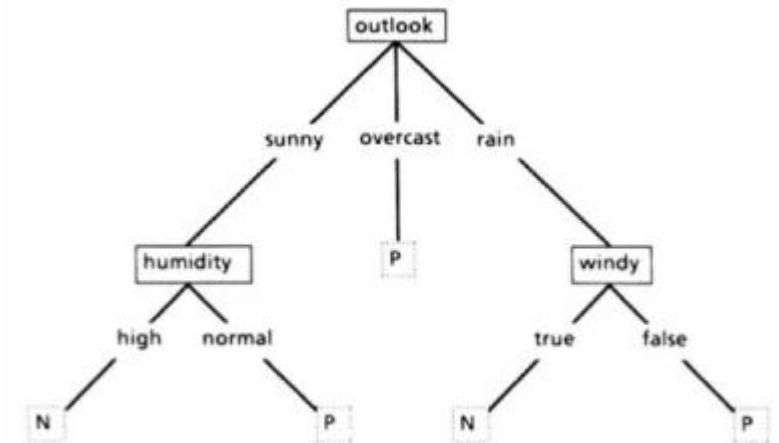
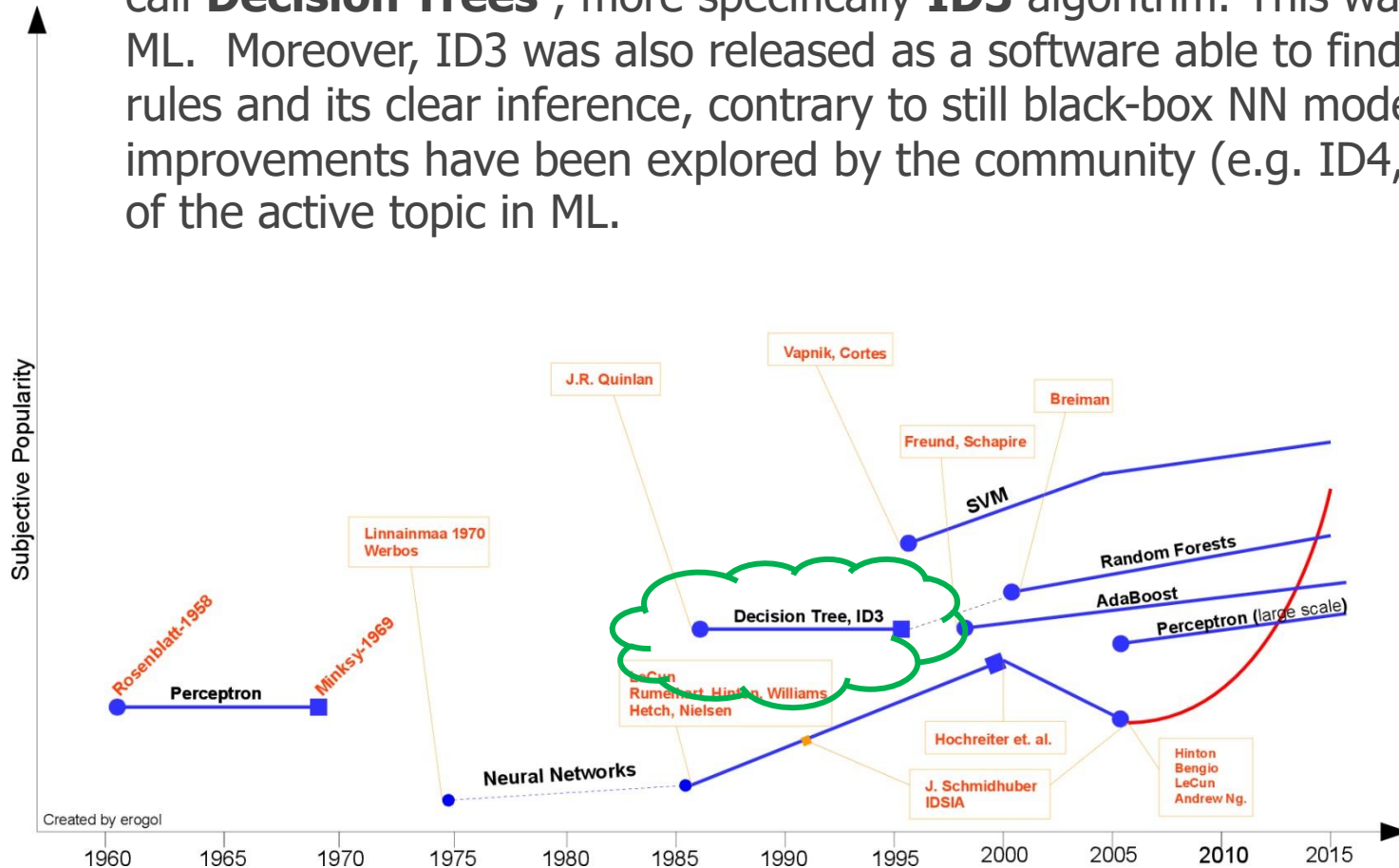


Figure 2. A simple decision tree

Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1.1 (1986): 81-106.

# History of Machine Learning (From Eren Golge's Blog)

One of the most important ML breakthrough was **Support Vector Machines** (Networks) (SVM), proposed by **Vapnik and Cortes** in **1995** with very strong theoretical standing and empirical results. That was the time separating the ML community into two crowds as NN or SVM advocates. However the competition between two community was not very easy for the NN side after **Kernelized** version of SVM by **near 2000s** .(I was not able to find the first paper about the topic), SVM got the best of many tasks that were occupied by NN models before. In addition, SVM was able to exploit all the profound knowledge of convex optimization, generalization margin theory and kernels against NN models. Therefore, it could find large push from different disciplines causing very rapid theoretical and practical improvements.

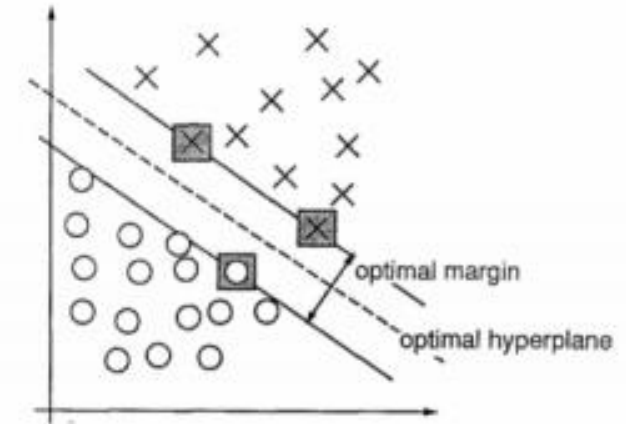
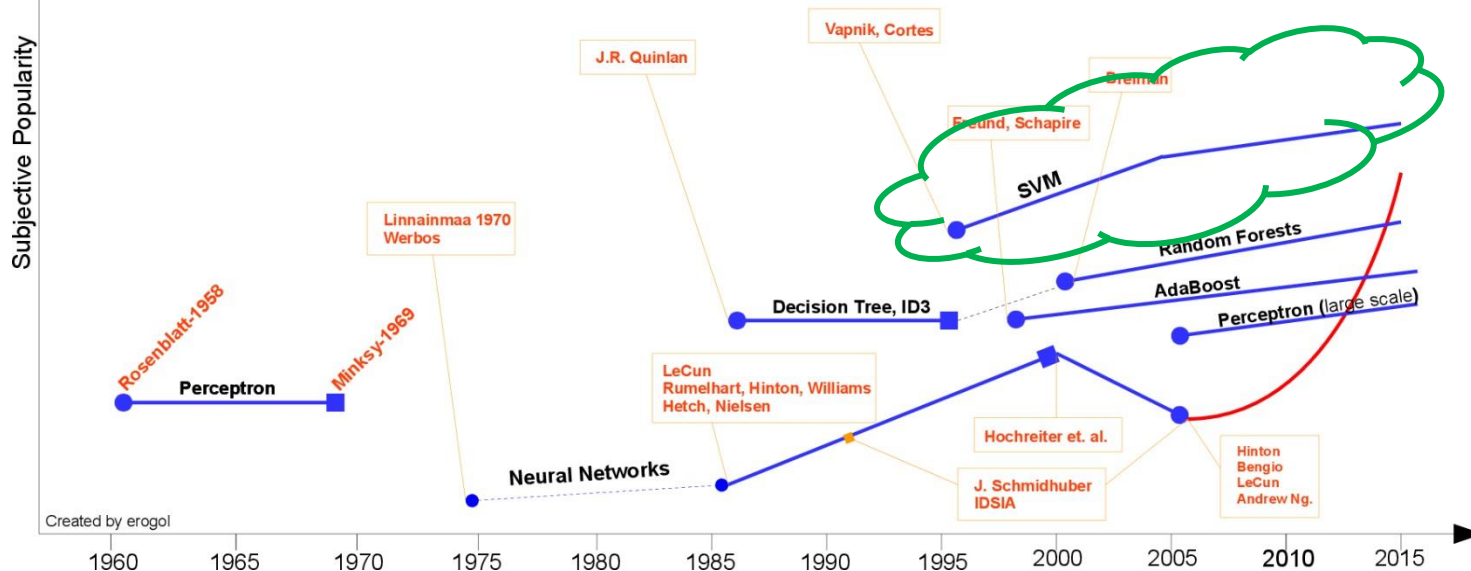
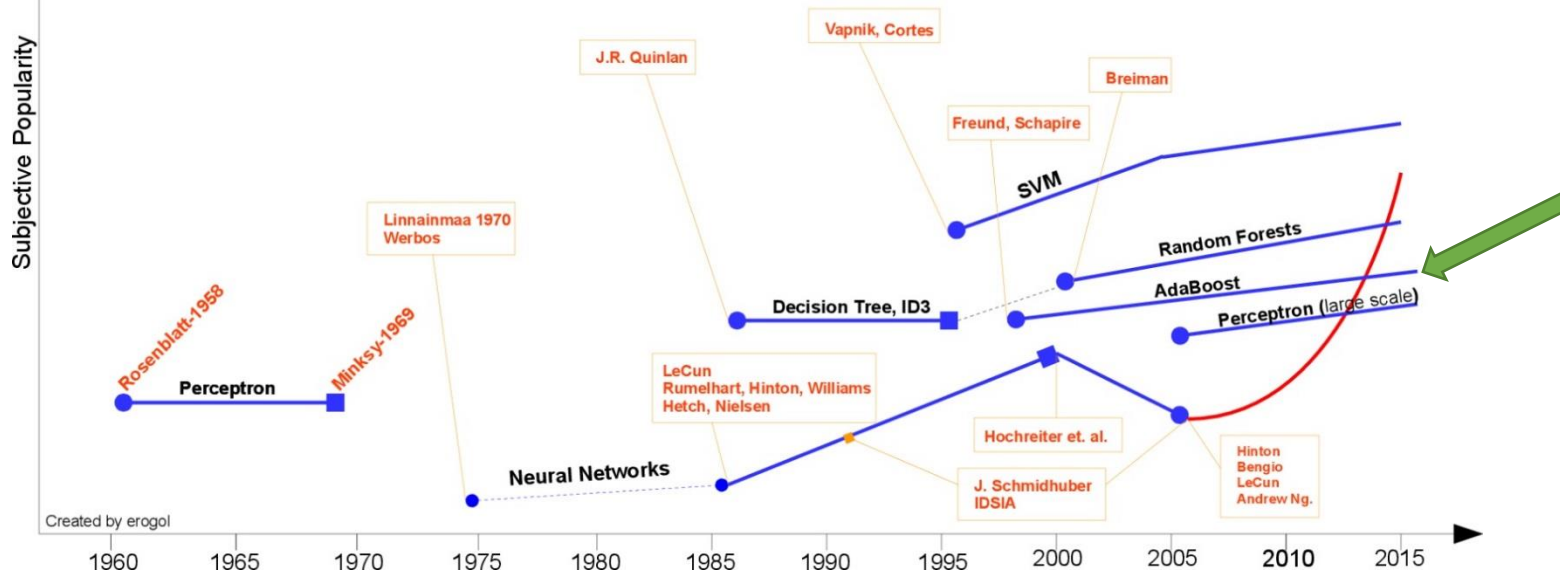


Figure 2. An example of a separable problem in a 2 dimensional space. The support vectors, marked with grey squares, define the margin of largest separation between the two classes.

# History of Machine Learning (From Eren Golge's Blog)

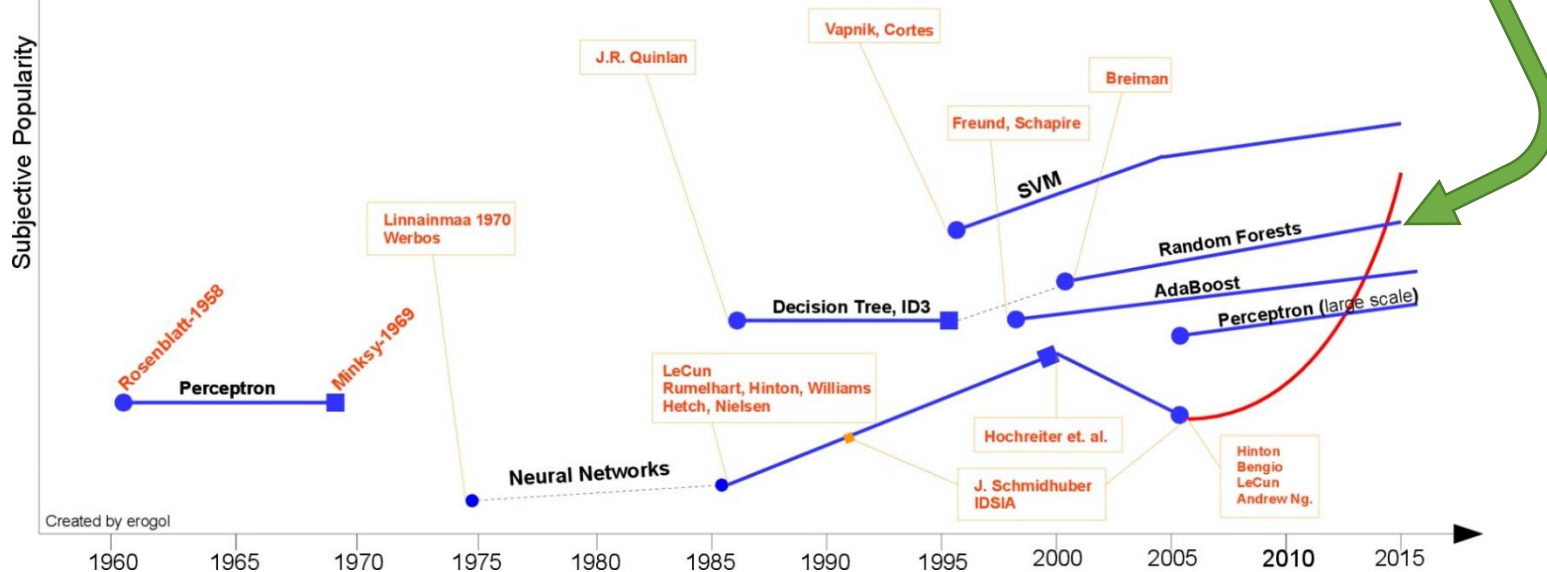
NN took another damage by the work of Hochreiter's thesis in 1991 and **Hochreiter et. al. in 2001**, showing the gradient loss after the saturation of NN units as we apply BP learning. Simply means, it is redundant to train NN units after a certain number of epochs owing to saturated units hence NNs are very inclined to over-fit in a short number of epochs.



Little before, another solid ML model was proposed by **Freund and Schapire** in **1997** prescribed with boosted ensemble of weak classifiers called **Adaboost**. This work also gave the Godel Prize to the authors at the time. Adaboost trains weak set of classifiers that are easy to train, by giving more importance to hard instances. This model still the basis of many different tasks like face recognition and detection.

# History of Machine Learning (From Eren Golge's Blog)

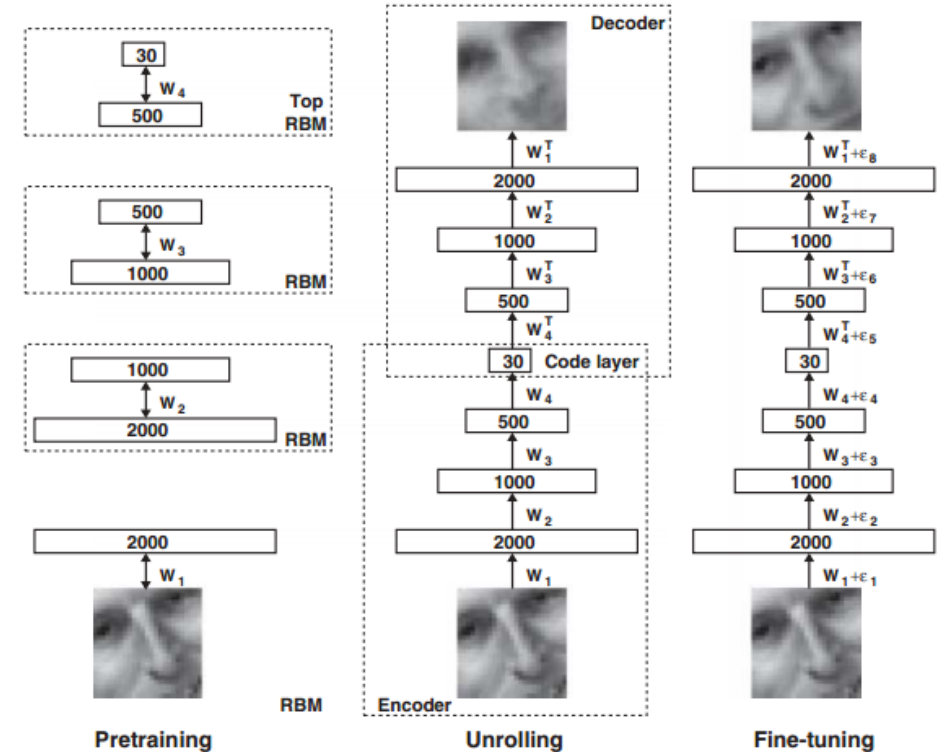
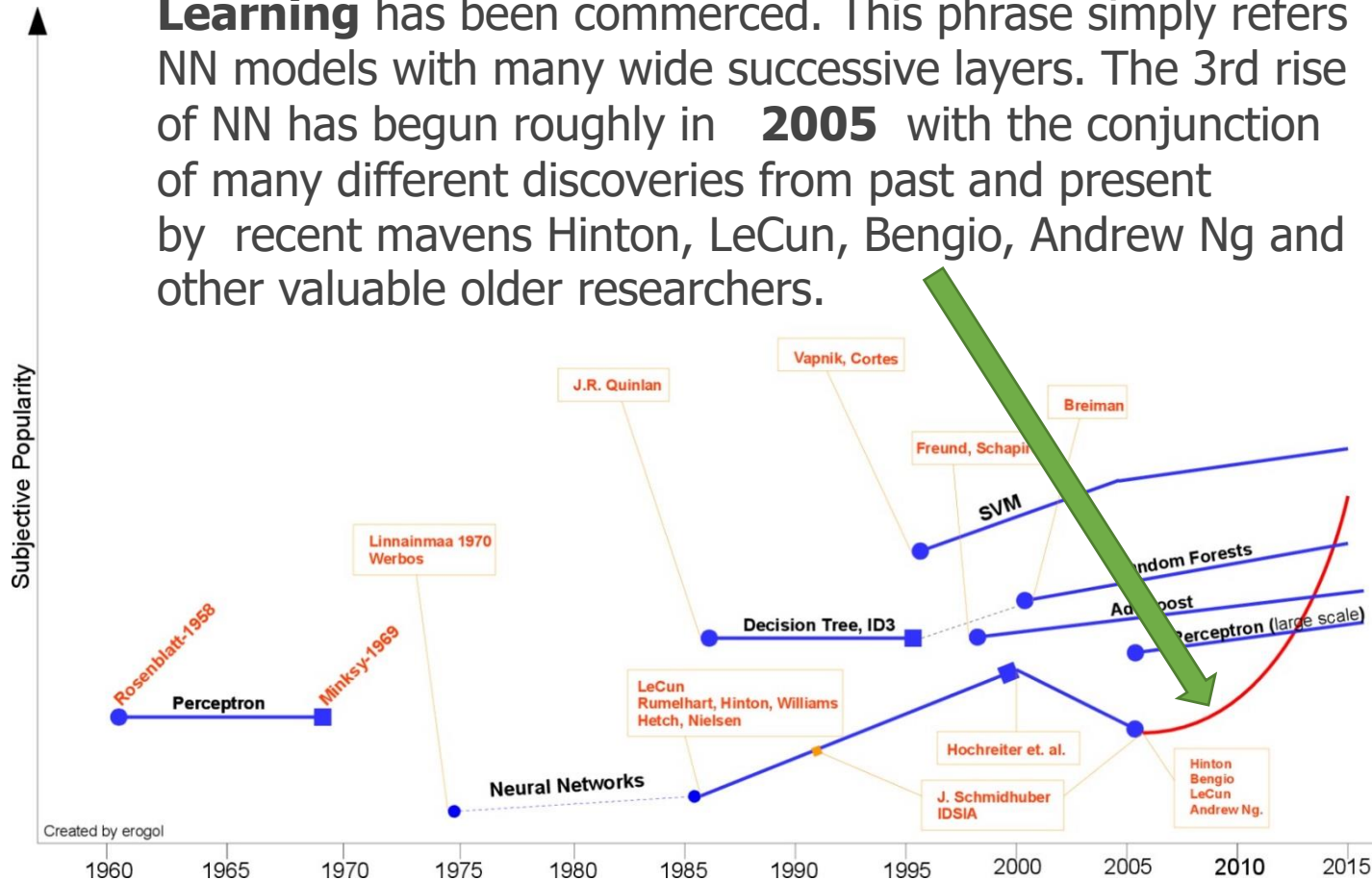
Another ensemble model explored by **Breiman** in **2001** that ensembles multiple decision trees where each of them is curated by a random subset of instances and each node is selected from a random subset of features. Owing to its nature, it is called **Random Forests(RF)**. RF has also theoretical and empirical proofs of endurance against over-fitting. Even AdaBoost shows weakness to over-fitting and outlier instances in the data, RF is more robust model against these caveats.(For more detail about RF, refer to my old post.). RF shows its success in many different tasks like Kaggle competitions as well.



Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large.

# History of Machine Learning (From Eren Golge's Blog)

As we come closer today, a new era of NN called **Deep Learning** has been commenced. This phrase simply refers NN models with many wide successive layers. The 3rd rise of NN has begun roughly in **2005** with the conjunction of many different discoveries from past and present by recent mavens Hinton, LeCun, Bengio, Andrew Ng and other valuable older researchers.



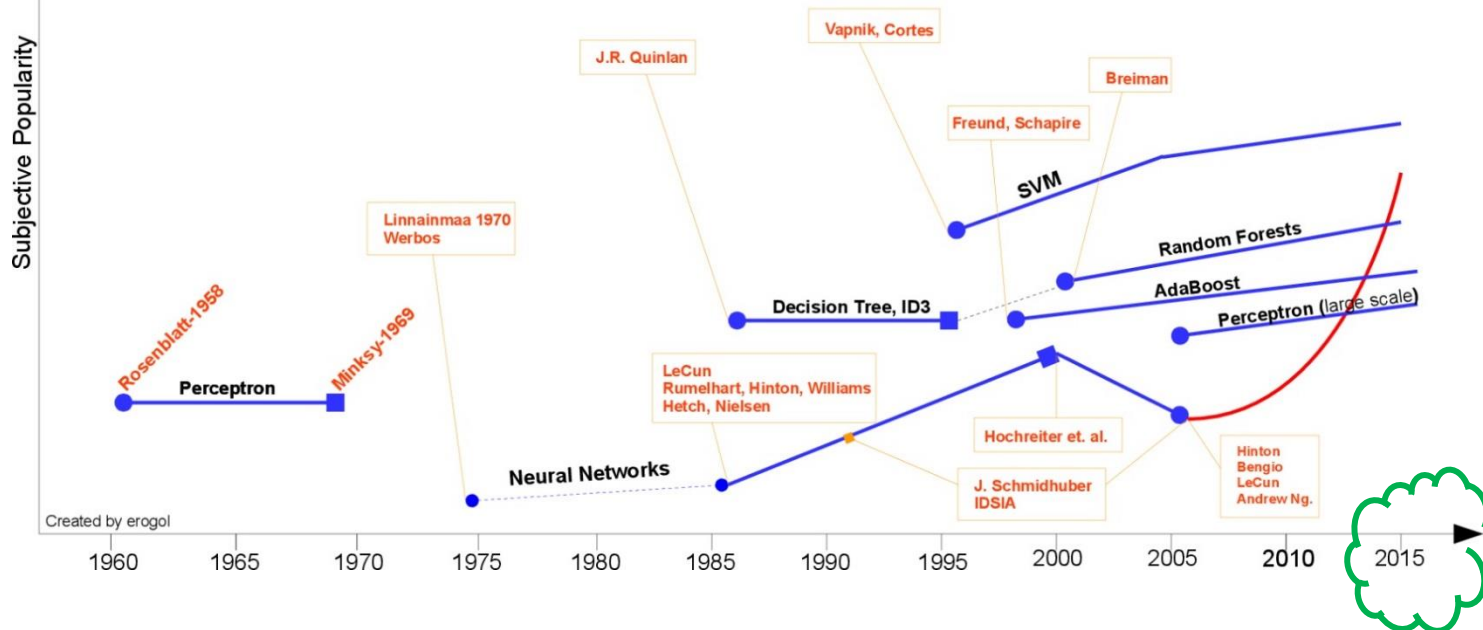
Hinton, G. E. and Salakhutdinov, R. R.  
Reducing the dimensionality of data with neural networks.  
Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006..



# History of Machine Learning

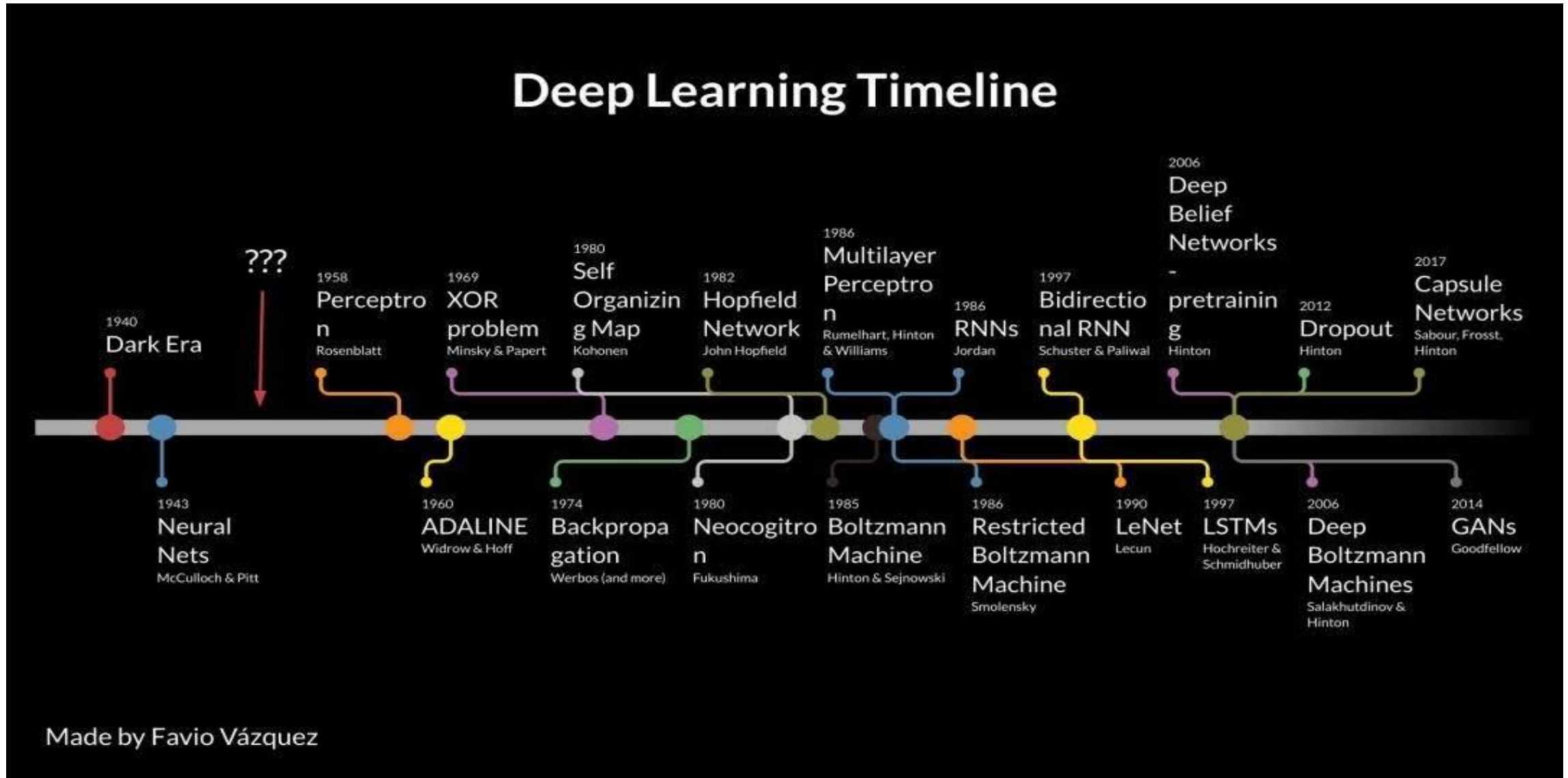
With the combination of all those ideas and non-listed ones, NN models are able to beat off state of art at very different tasks such as Object Recognition, Speech Recognition, NLP etc.

However, it should be noted that this absolutely does not mean, it is the end of other ML streams. Even Deep Learning success stories grow rapidly, there are many critics directed to training cost and tuning exogenous parameters of these models.

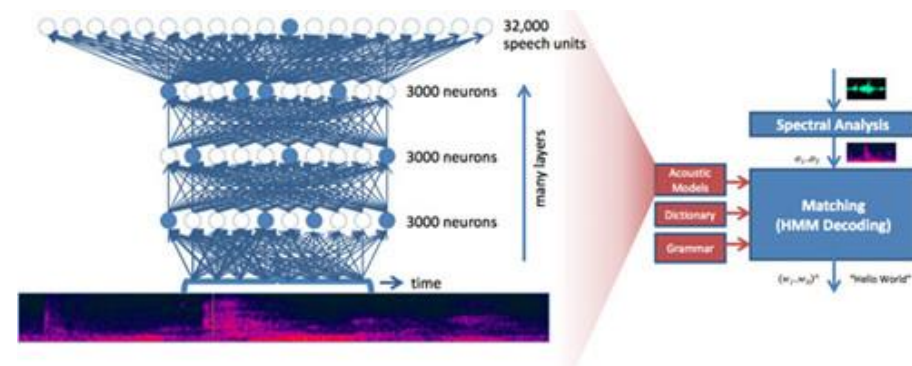


After the growth of WWW and Social Media, a new term, **BigData** emerged and affected ML research wildly. Because of the large problems arising from BigData, many strong ML algorithms are useless for reasonable systems (not for giant Tech Companies of course). Hence, research people come up with a new set of simple models that are dubbed **Bandit Algorithms** (formally predicated with **Online Learning**) that makes learning easier and adaptable for large scale problems.

# History of Deep Learning

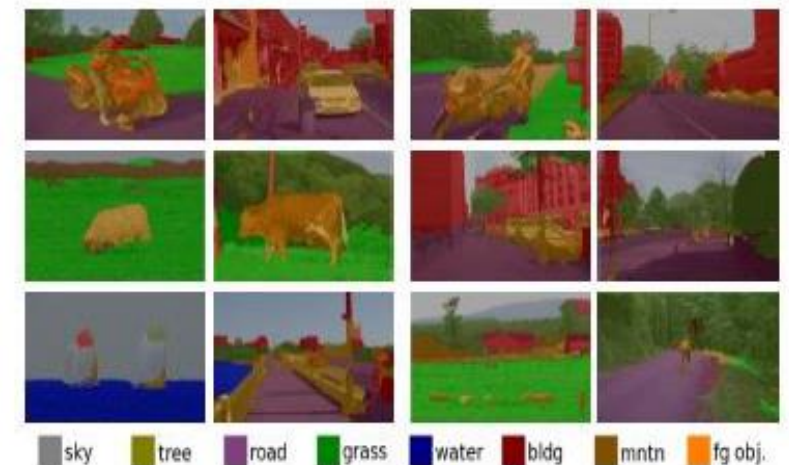
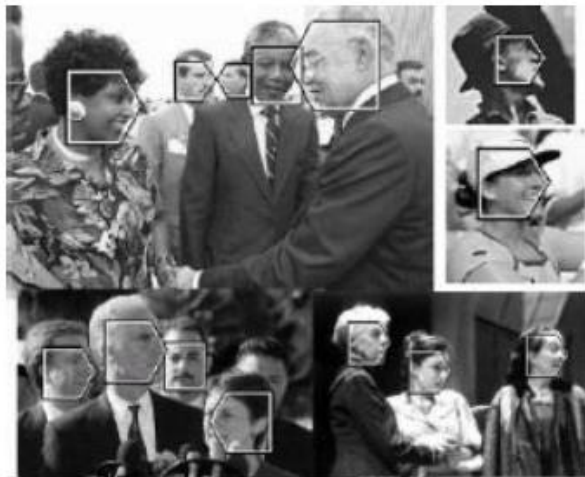


# Applications: Speech Recognition

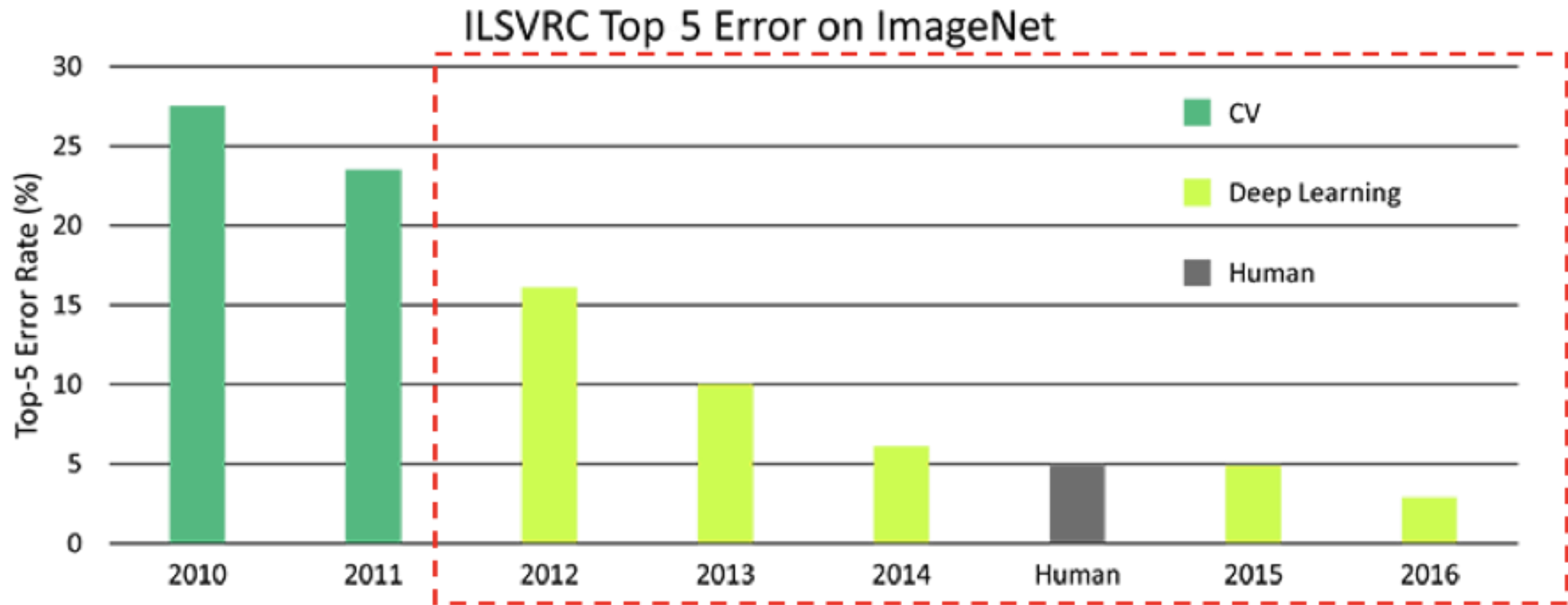




# Applications: Computer Vision



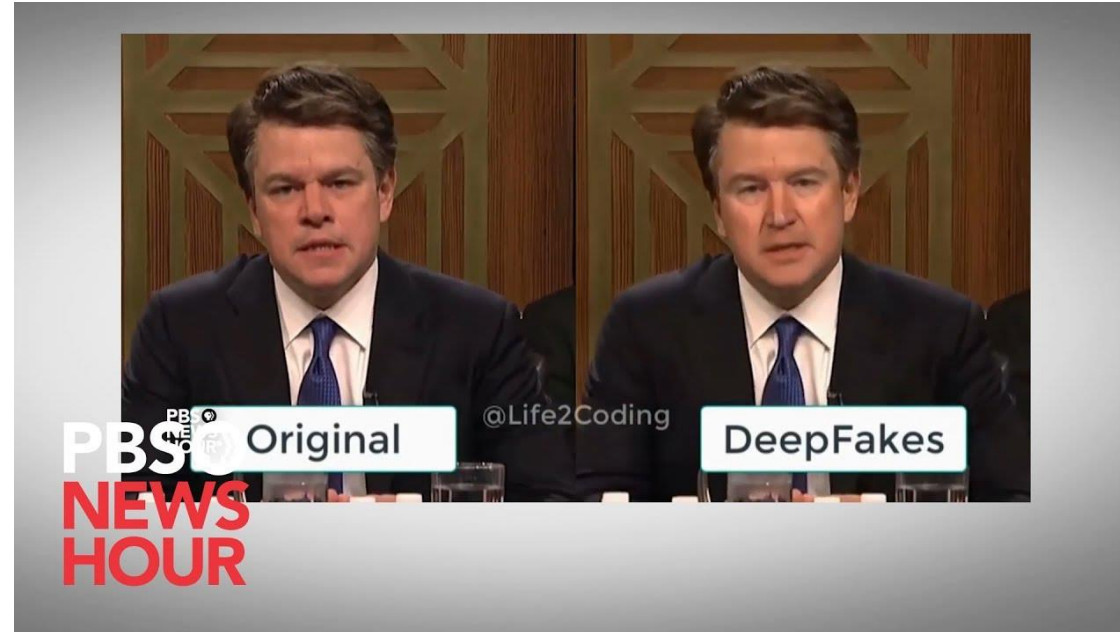
# Success of Deep Learning: Computer Vision



The introduction of Deep Learning techniques drove performance on image categorization from 30% error rates in 2010, down to <2% in 2017



# CV: Deefake



**Deepfake** is a technique for human image synthesis based on artificial intelligence.

Combine and superimpose existing images and videos onto source images or videos using a machine learning technique known as generative adversarial network.

# Applications: Web Search



**Baidu 百度** 中国科学院大学  百度一下

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约709,000个  搜索工具

[首页-中国科学院大学](#) [官网](#)

 **中国科学院大学** (英文名: University of Chinese Academy of Sciences) 成立于1978年, 是经国务院批准创办的第一所研究生院, 是国家教育部正式批准成立的正规高校, 是一所...  
[www.gucas.ac.cn/](http://www.gucas.ac.cn/) - 百度快照 - 94%好评

[中国科学院大学\\_百度百科](#)

 **中国科学院大学** (University of Chinese Academy of Sciences), 简称“国科大”, 中国科学技术最高学府, 隶属于中国科学院, 前身是中国科学院研究生院, 创建于1978年, 是经国务院批准创办的新中国第一所研究生院, 培养了中国的第一个理学...  
[历史沿革](#) [办学条件](#) [学术研究](#) [文化传统](#) [校园环境](#) [更多>>](#)  
[baike.baidu.com/](http://baike.baidu.com/)

[中国科学院大学吧\\_百度贴吧](#)

 博学笃志, 格物明德。  
关注用户: 6975人  
累计发帖: 3万  
[规定说明](#) [果壳纵览](#) [资源集勘](#) [人生之路](#)

[国科大学子自己的BT资源站——FBI2.0正式发布](#) 点击: 2万 回复: 86  
[《在前行中感悟, 在感悟中成长》国科大本科第一...](#) 点击: 1万 回复: 64  
[分享昨天撸的一组玉泉路校区照片](#) 点击: 2935 回复: 11  
[查看更多中国科学院大学吧的内容>>](#)  
[tieba.baidu.com/中国科学院大学?fr=ala0](http://tieba.baidu.com/中国科学院大学?fr=ala0)

[在北京市搜索中国科学院大学\\_百度地图](#)

以下结果已根据您的位置优化:

A [中国科学院大学雁栖湖校区](#) [全景](#)  
地址: 北京市怀柔区怀北镇怀北庄中国...  
电话: 010-69671021

B [中国科学院大学](#) [全景](#)  
★★★★★ 5条评论  
地址: 北京市石景山区玉泉路19号甲(...  
电话: 010-88256030

C [中国科学院大学中关村园区](#) [全景](#)

**相关人物** [展开](#)

 **吴天湛**  
北京大学优秀志愿者

 **冯翔**  
中国科学院博士

 **李小文**  
重庆邮电大学研究员

 **丁仲礼**  
第十届全国政协委员

**相关院校** [展开](#)

 **福建理工大学**  
副部级高校

 **吉林理工大学**  
顶尖军工学校

 **陆军大学**  
军事学府一所最高级别

 **国防科学技术大学**  
全国重点大学

**学子向往的211工程大学** [展开](#)

 **清华大学**  
科技英才的摇篮

 **清华大学**  
红色工程师的摇篮

 **武汉大学**  
综合性研究型重点大学

 **南京大学**  
中国科学社的大本营

**高校风云榜**

排名	搜索指数
1 <b>清华大学</b>	5638 ↓
2 <b>北京大学</b>	4700 ↓





# Applications: Recommender System

The screenshot shows the Amazon.cn homepage. At the top, there's a navigation bar with the Amazon logo, a search bar, and links like '全部分类' (All Categories), '我的亚马逊' (My Amazon), 'Z秒杀' (Z秒杀), '礼品卡' (Gift Card), '我要开店' (I want to open a store), '海外购' (Overseas Purchase), '帮助' (Help), and 'In English'. Below the navigation bar, there's a large promotional banner for books with the text '好书 任选 15万畅销书 满100减20' (Good books, choose any, 150,000 best-selling books, full 100 off 20). To the left of the banner is a sidebar with a list of product categories: Kindle电子书阅读器, Fire平板电脑, 亚马逊海外购, 图书, 音像、软件, 手机、摄影、数码, 家居、厨具、家装, 电脑、办公, 家用电器, 美妆、个护健康, 食品、酒水、生鲜, 玩具、母婴、家庭会员, 运动户外, 服装、鞋靴、箱包, 钟表、首饰, 汽车用品, 亚马逊 Cloud Drive, and 全部商品分类. Below the banner, there are two promotional tiles. The left tile is for '今日 镇店之宝' (Today's Store Treasure) featuring a couple and a price of ¥59.00. The right tile is for '海外购 直采' (Overseas Purchase Direct Procurement) featuring a jar of Kirkland Unsalted Mixed Nuts and a price of ¥180.00. At the bottom, there's a section titled '☆ 与您浏览过的商品相关的推荐' (Recommendations related to the products you have browsed) showing a row of hair styling tools like flat irons and curlers.



The screenshot shows the Netflix 'House of Cards' page. On the left is a poster for the show featuring Kevin Spacey and Robin Wright. On the right is the text: 'Watch Season 3 Now', 'This Emmy-winning original thriller series stars Golden Globe winner Kevin Spacey as ruthless, cunning Congressman Francis Underwood, who will stop at nothing to conquer the halls of power in Washington D.C. His secret weapon: his gorgeous, ambitious, and equally conniving wife Claire (Golden Globe winner Robin Wright).', 'Starring: Kevin Spacey, Robin Wright, Kate Mara', 'Genres: TV Shows, TV Dramas', and 'This show is: WB, Crime, Dark'.

# Applications: Social Computing





# Applications: Natural Language Processing



Machine translation

Original text  
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。

Phrase-based translation	Neural machine translation	Human translation
Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang <b>will</b> start the annual dialogue mechanism with Prime Minister Trudeau <b>of Canada and hold the first annual dialogue between the two premiers.</b>	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

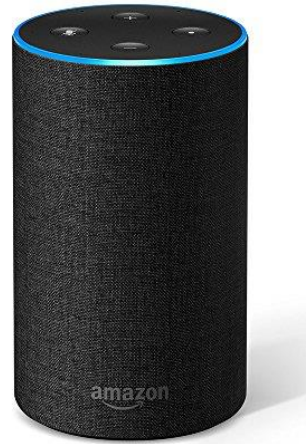
Google's neural machine translation reduced translation errors by an average of 60% when compared to the prior google translate technology.



# Applications: Natural Language Processing



Siri



Alexa

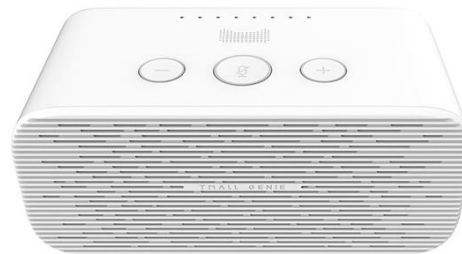


Hey Cortana

Cortana



Xiaomi



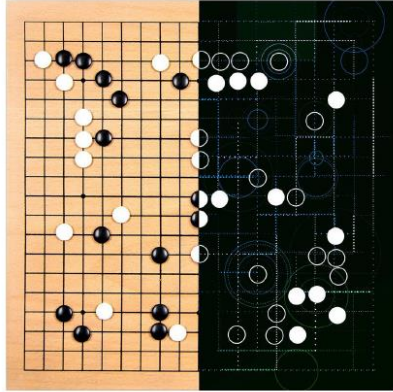
Tmall



Google Assistant

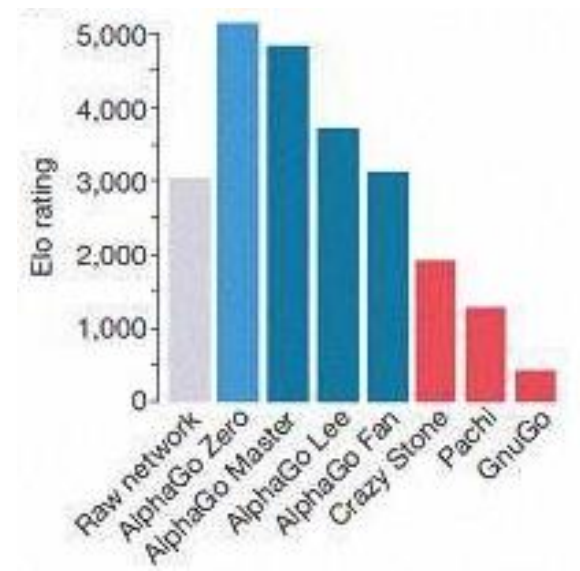
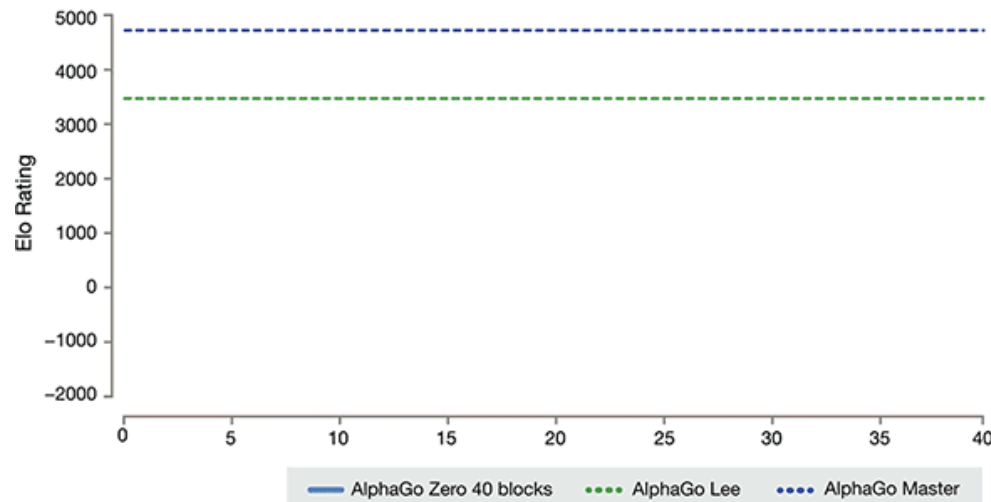
Google Assistant

# Applications: Game



# TimeLine of AlphaGo

- 2016.1.27 AlphaGo Fan vs fan hui : 5-0
- 2016.3.9 – 2016.3.15 AlphaGo Lee vs li shishi : 4-1
- 2017.5.23 – 2017.5.27 AlphaGo Master vs ke jie : 3-0
- Now AlphaGo Zero





# Applications: Game



- 2010: the first AI competition.
- 2012: The use of persistent file storage enables the Bot to learn throughout the competition. Bots have emerged that can predict results.
- 2016: studying starcraft AI is no longer about beating human players, but about how it makes decisions in more complex environments.
- 2019: AlphaStar vs TLO, Mana: 10-1

# More Applications





# Human Society Affairs







5 min Break

