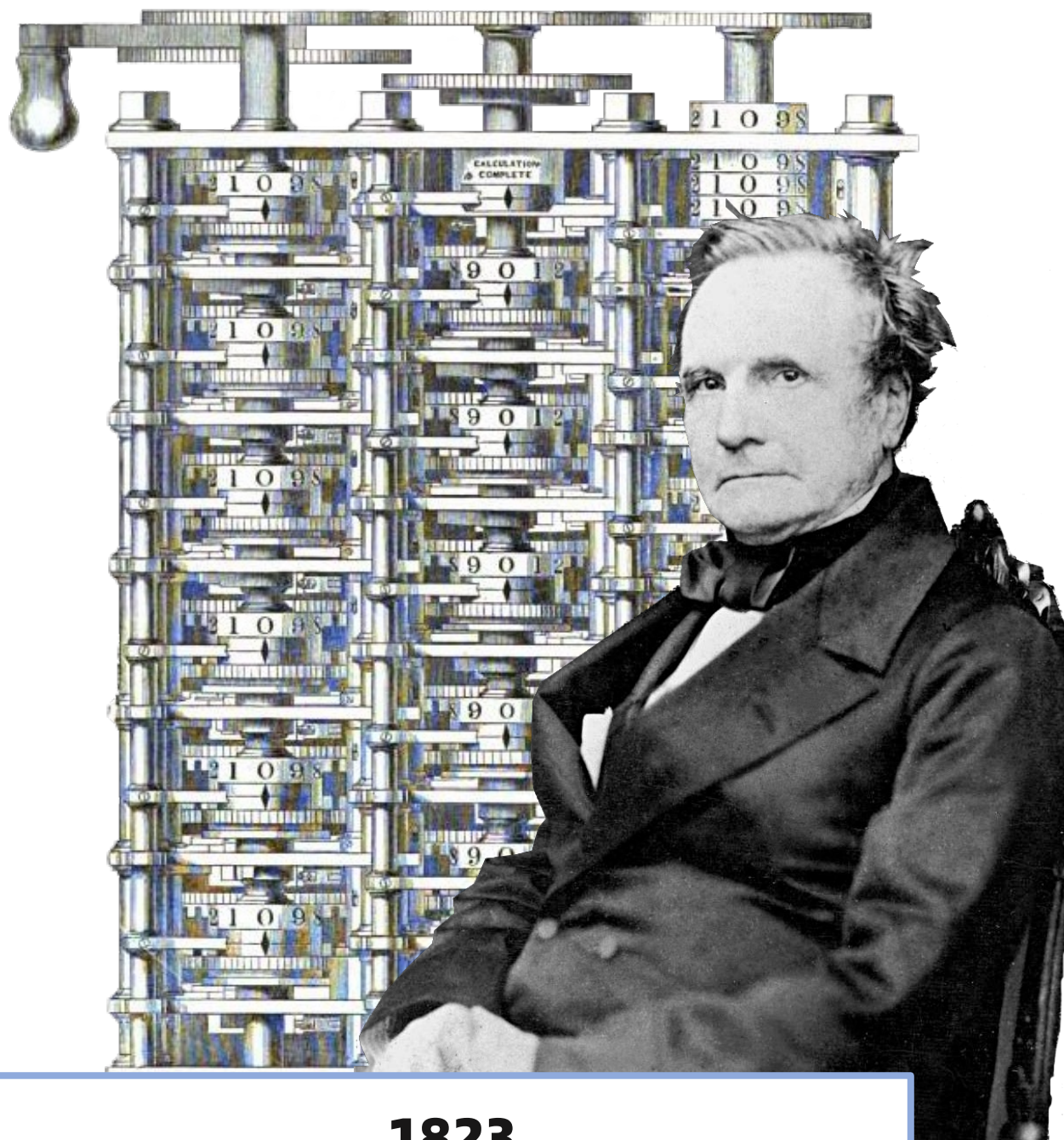




全图说机器学习技术发展历史

人工智能是一个庞大的课题，文本选择最重要的机器学习（Machine Learning）视角为锚点，从科学技术的发展维度切入，尝试梳理从人类早期信息论、控制论，到神经网络、以及各种机器学习算法的革新和迭代中的各种线索。文本尝试带着你拨开迷雾，把机器学习发展路径中的关键的科学家/历史事件总结出来，按照关键事件发生的年代顺序排列和线性展开，试图说明白机器学习技术发展的前世今生。希望对大家有所帮助和启迪。

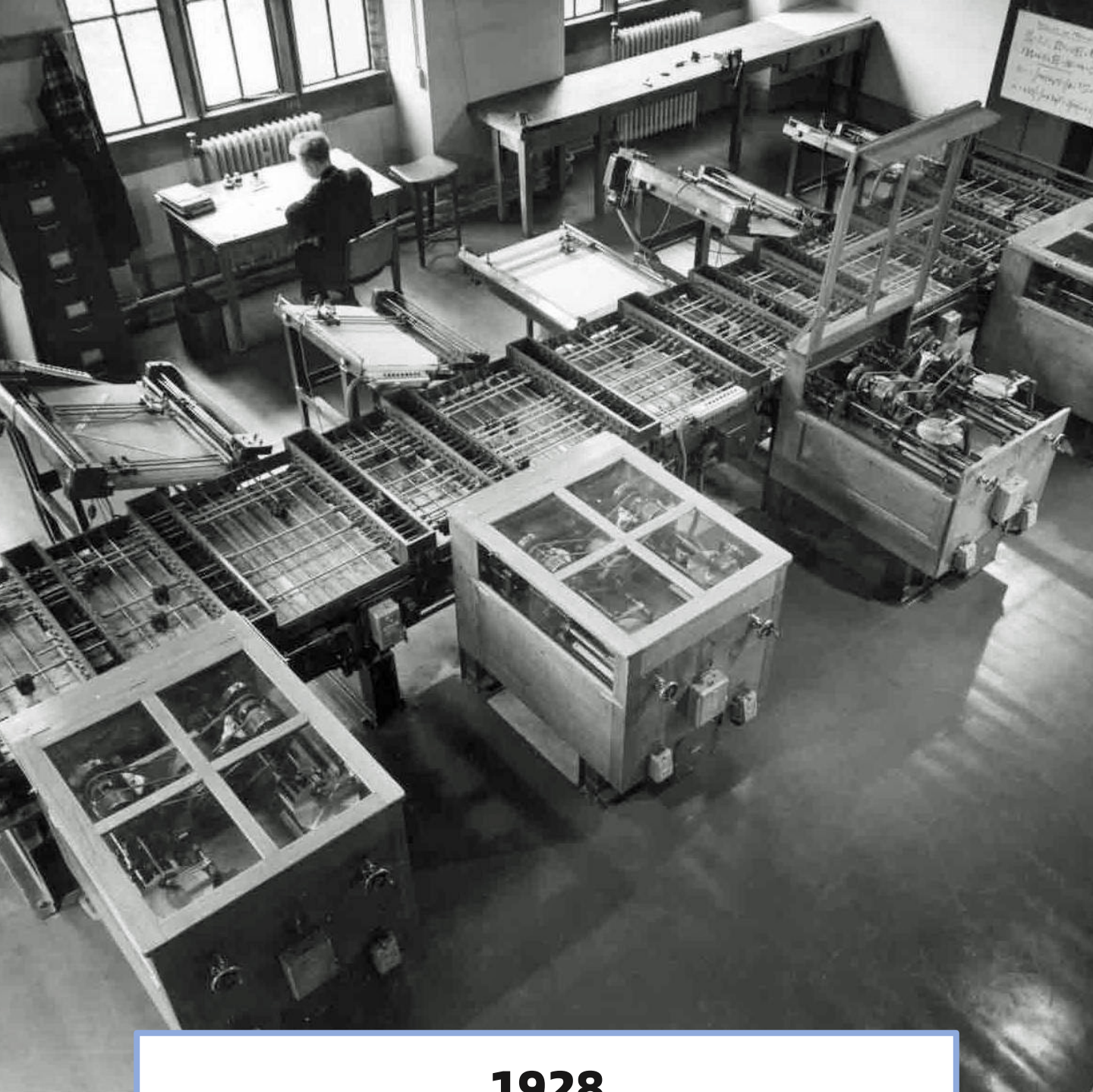




1823

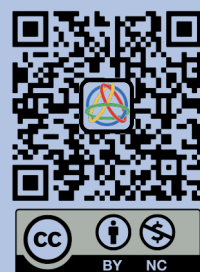
查尔斯·巴贝奇提出了可编程计算机的概念。被认为是“计算机之父”，在19世纪初构思并发明了第一台机械计算机。1823年，启动分析仪（analytical engine）设计开发，有存储和运算部件。但并未建成。艾达·洛夫莱斯（Ada Lovelace）推测该机器“可能会创作出任何复杂程度或范围的精心制作和科学的音乐作品”。





1928

首个广泛实用的通用微分分析仪是由Vannevar Bush在1928-1931年在麻省理工学院建造。他在1931年发表了一篇关于该设备的进一步文章时称其为“差分分析仪”。被誉为“思考机器”和“机械大脑”。

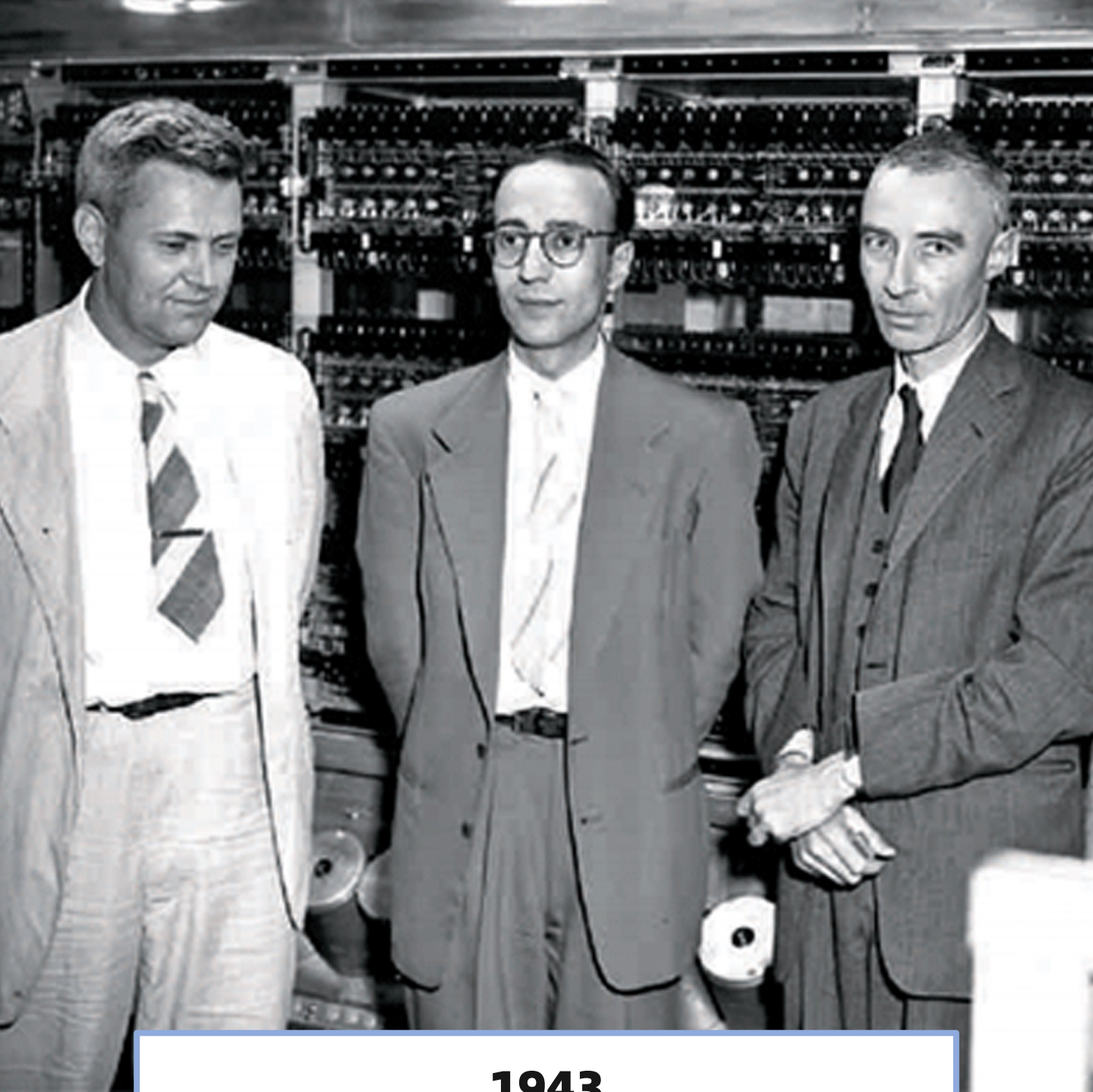




1941

1941 年，康拉德·楚泽发明第一台可编程机电计算机 Z3。这被称为世界上第一个有图灵完全性程控功能的图灵计算机。

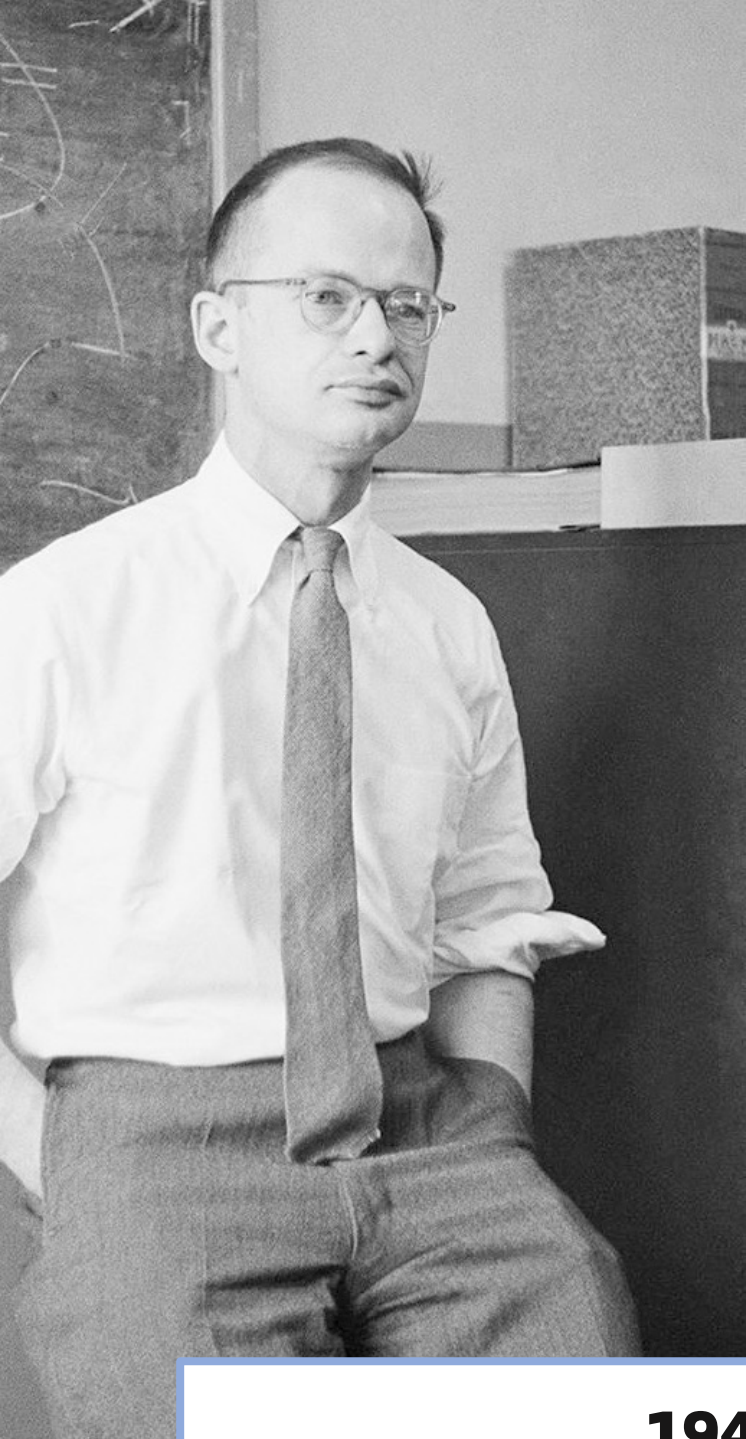




1943

1943年，著名的控制论（cybernetics）和现代目的论的奠基论文之一，“行为、目的和目的论”在科学哲学上发表。阿图罗·罗森布鲁斯（Arturo Rosenblueth）、维纳（Norbert Wiener）和朱利安·毕格罗（Julian Bigelow）作为主要作者，成为控制论的先驱。

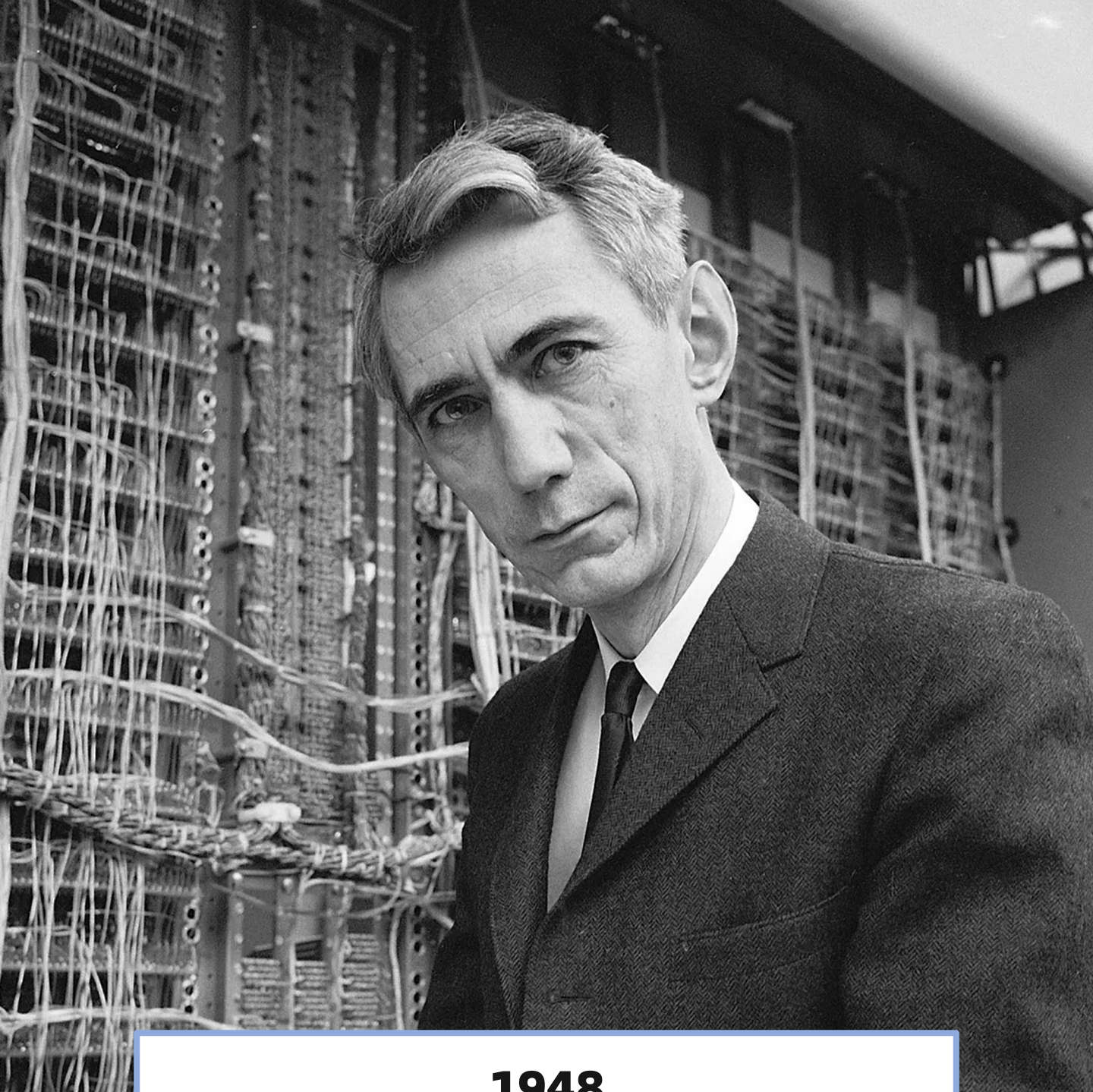




1943

沃伦·麦卡洛克（Warren McCulloch）和沃尔特·皮茨（Walter Pitts）在1943年一起撰写了一篇科学史上的开创性论文“神经活动中内在思想的逻辑演算”。文中提出了第一个神经网络的数学模型。该模型的单元，一个简单的形式化神经元，至今仍是神经网络领域的参考标准。它通常被称为McCulloch-Pitts 神经元。

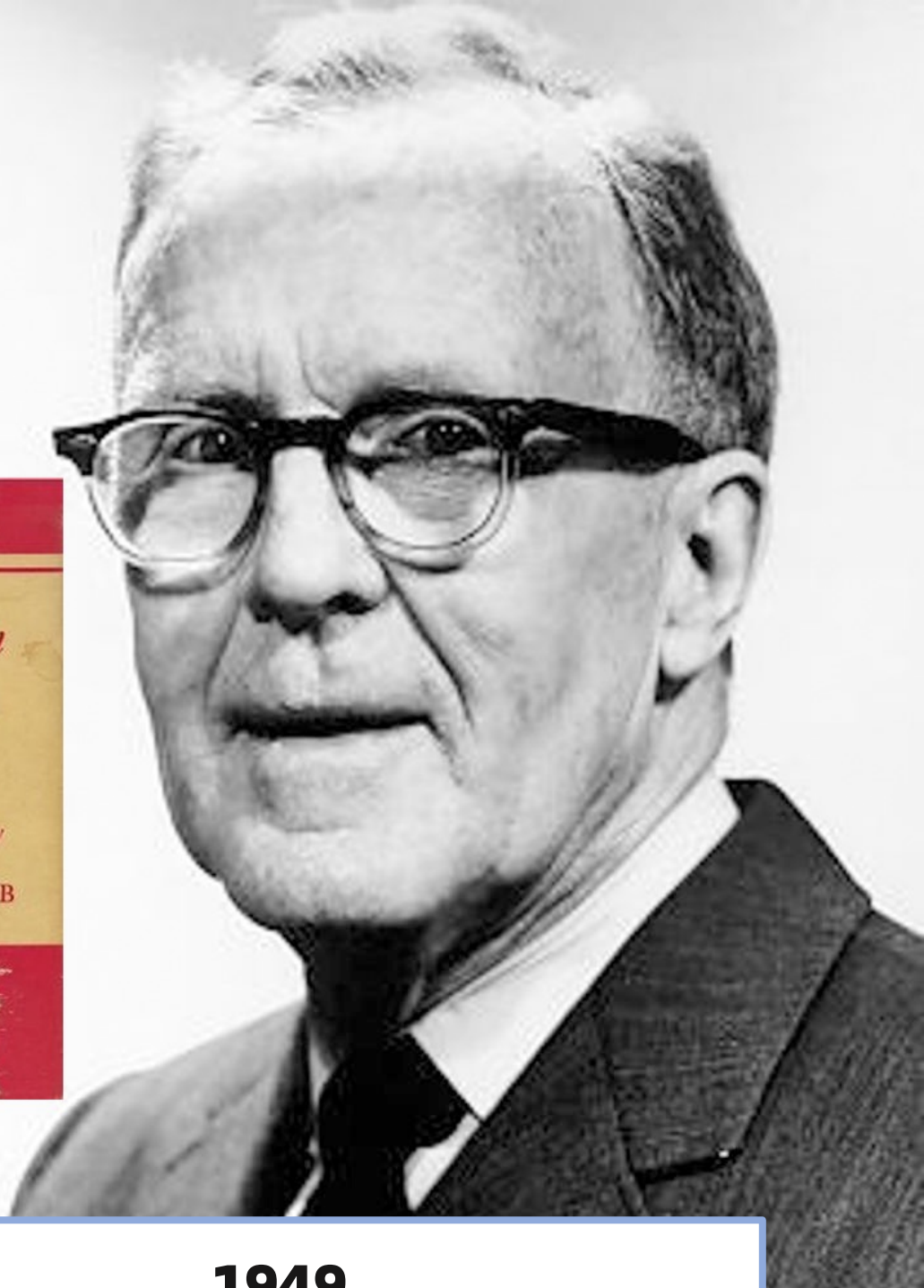
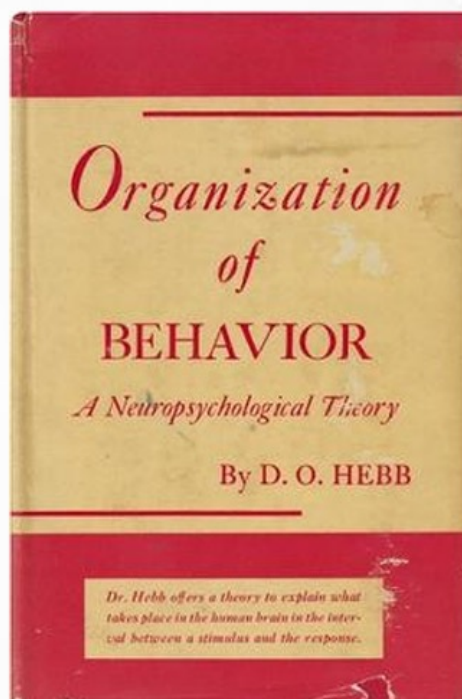




1948

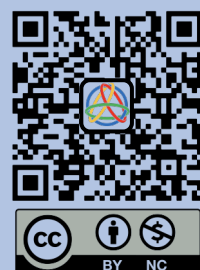
1948 年，克劳德·香农（Claude Shannon）发表了《通信的数学理论》，提出信息论。信息论是自然语言处理的基础。因此他被誉为“信息论之父”。

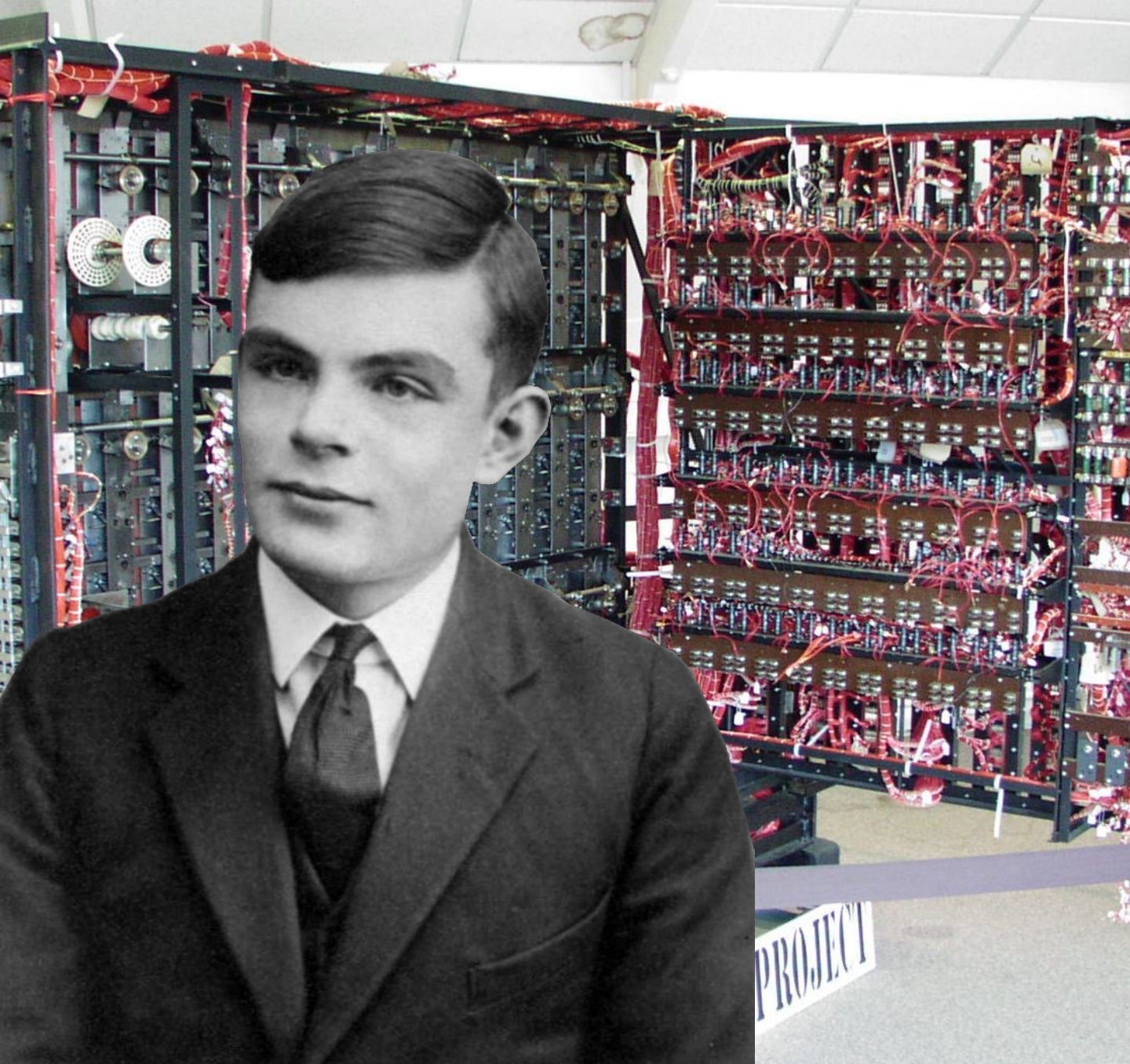




1949

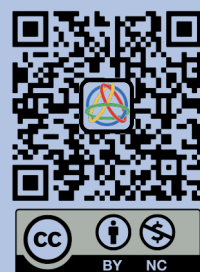
唐纳德·赫布（Donald Hebb）在 1949 年出版了经典著作《行为组织》。提出了神经元的功能如何促进学习的过程，其中最著名的是赫布学习理论。被誉为神经心理学和神经网络之父。赫布学习是无监督学习。后续演变成长期增强的模型。





1950

1950 年，艾伦图灵发表了一篇具有里程碑意义的论文，他在论文中推测了创造会思考的机器的可能性。并设计了他著名的图灵测试，一个测试人工智能是否达到人类智能水准的实验。图灵测试是人工智能哲学中第一个严肃的提议。



1956 Dartmouth Conference: The Founding Fathers of AI



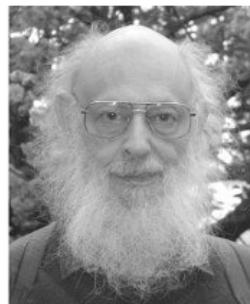
John MacCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



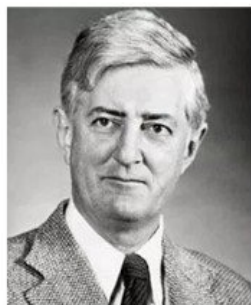
Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



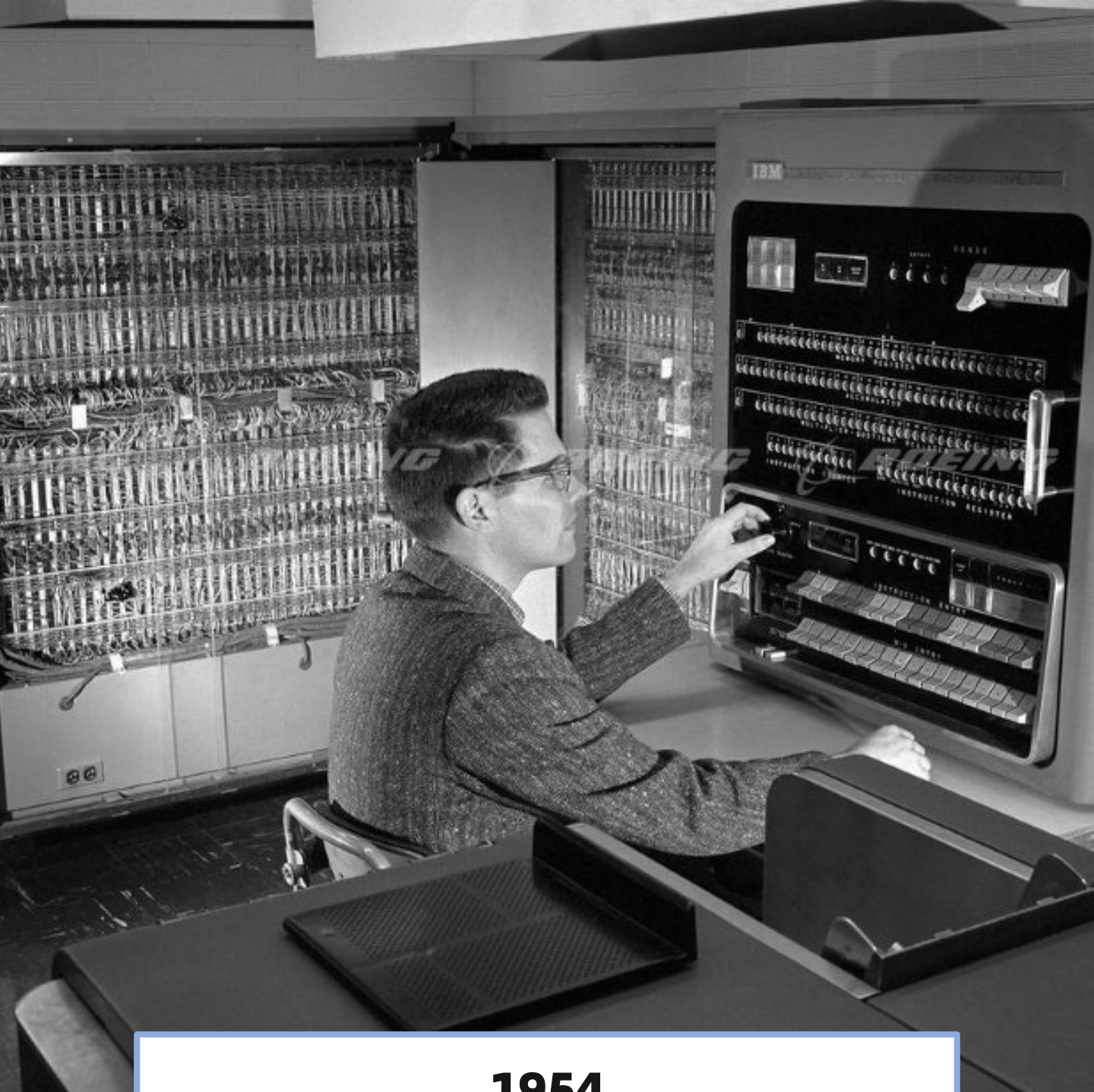
Trenchard More

1956

1956 年的达特茅斯研讨会，提案包括这样的断言：

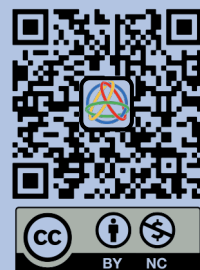
“学习的每个方面或智能的任何其他特征都可以如此精确地描述，以至于可以制造一台机器来模拟它”。会上约翰·麦卡锡（John McCarthy）说服与会者接受“人工智能”作为该领域的名称。这次会议被称为是人工智能诞生的会议。

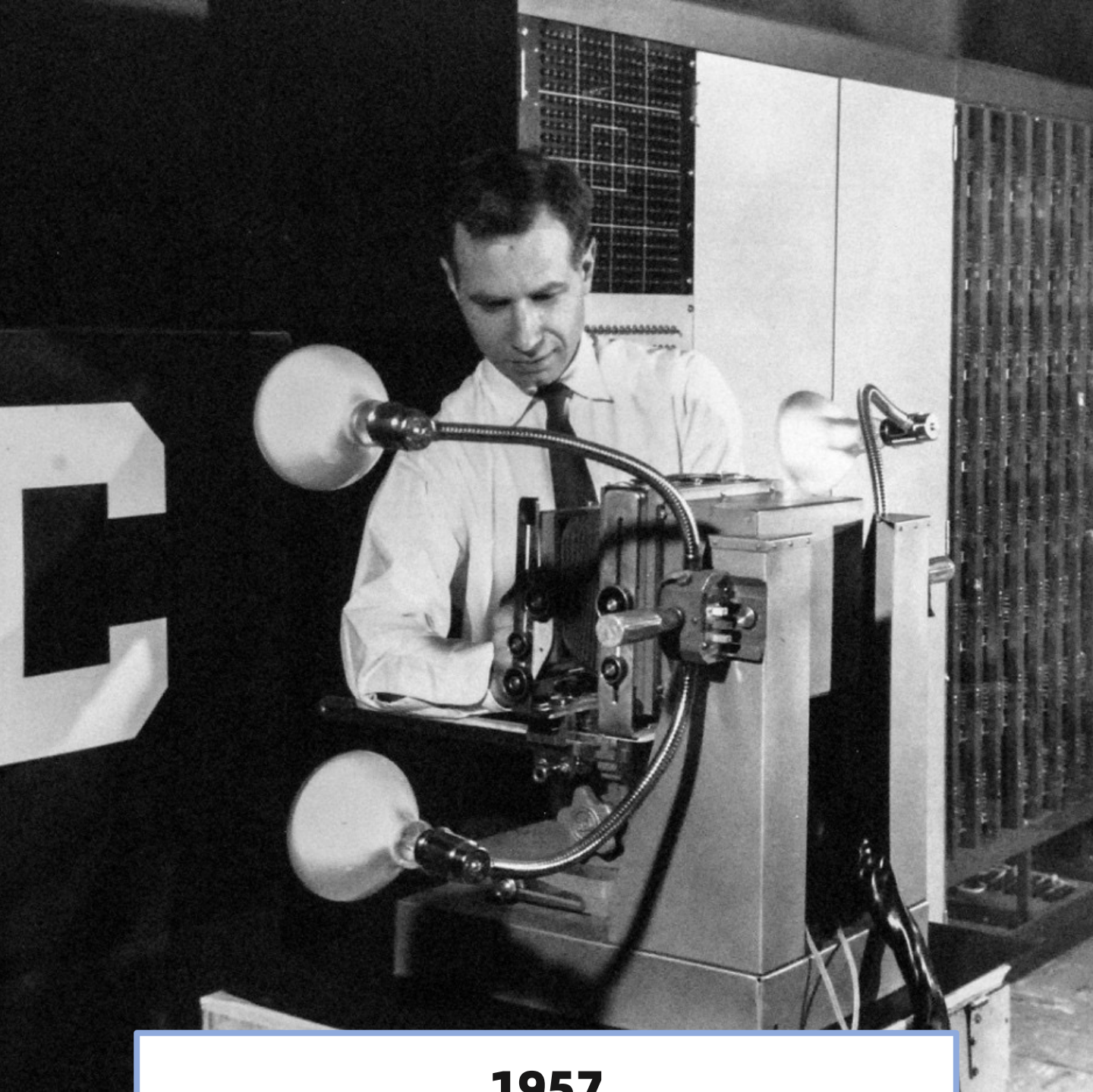




1954

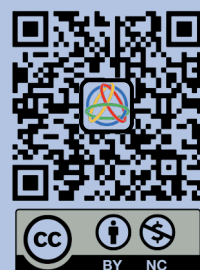
乔治敦-IBM 实验，是1954年1月7日进行的一次机器翻译演示。这个实验将60 多个俄语句子完全自动翻译成英语。但是由于低估了自然语言处理（NLP）的复杂性，实验进展缓慢，直到80年代，随着机器学习算法的改进，才改变了状况。





1957

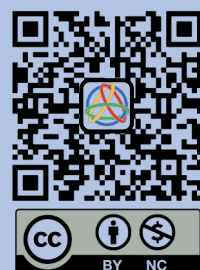
美国著名的人工智能领域心理学家弗兰克·罗森布拉特（Frank Rosenblatt）于1957年，在康奈尔航空实验室的IBM 704计算机上开发了一个神经计算机的Mark I 感知器，这是最早期的图像识别机器，但是它的视觉识别能力很弱，导致神经网络研究受挫和停滞。但是弗兰克还是被称为深度学习之父。





1959

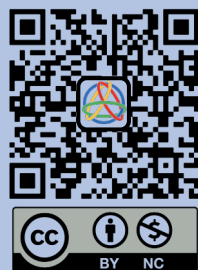
亚瑟·李·塞缪尔（Arthur Lee Samuel）创建了最早的软件哈希表而受到赞誉。他在1959年一台商用计算机 IBM 701上制作了第一个西洋跳棋程序。塞缪尔跳棋程序是世界上第一个成功的自学程序之一。他在1952年首次提出了“机器学习”一词。是美国计算机游戏和人工智能领域的先驱。

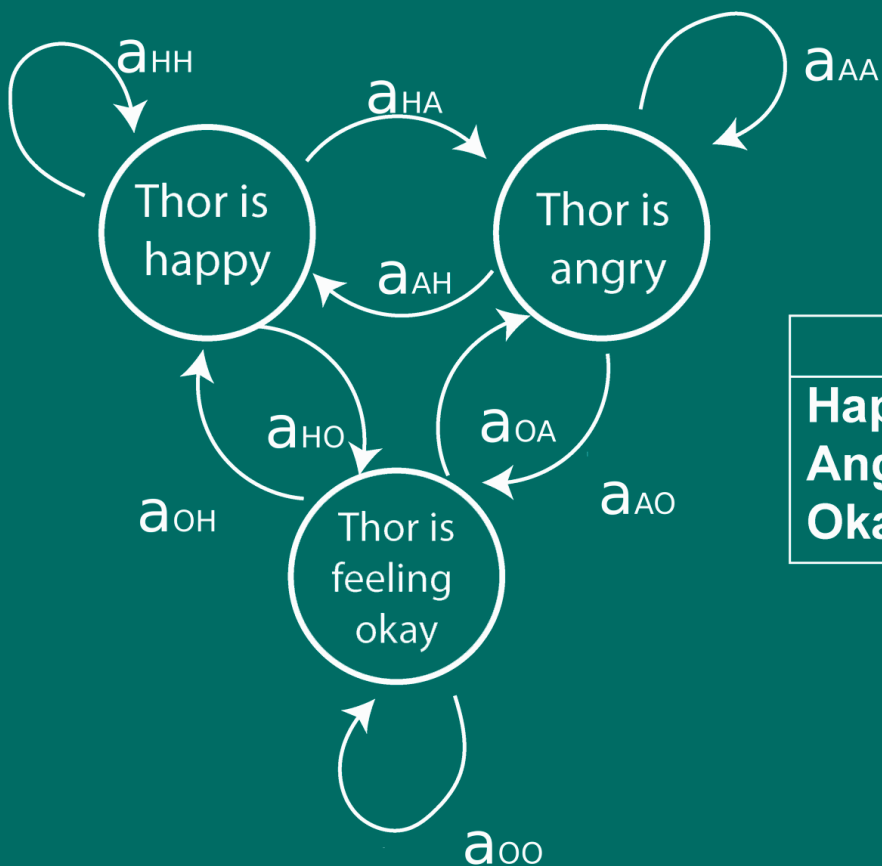




1960

鲁道夫·卡尔曼 (Rudolf Kálmán) 在1960年和理查德·比西 (Richard S. Bucy) 一起，发明卡尔曼滤波器（或 Kalman-Bucy 滤波器）。卡尔曼滤波器广泛应用于信号处理和计量经济学等领域。也用在机器人运动规划和控制中。

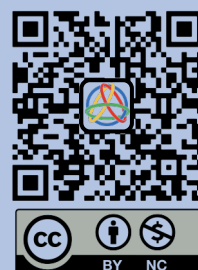


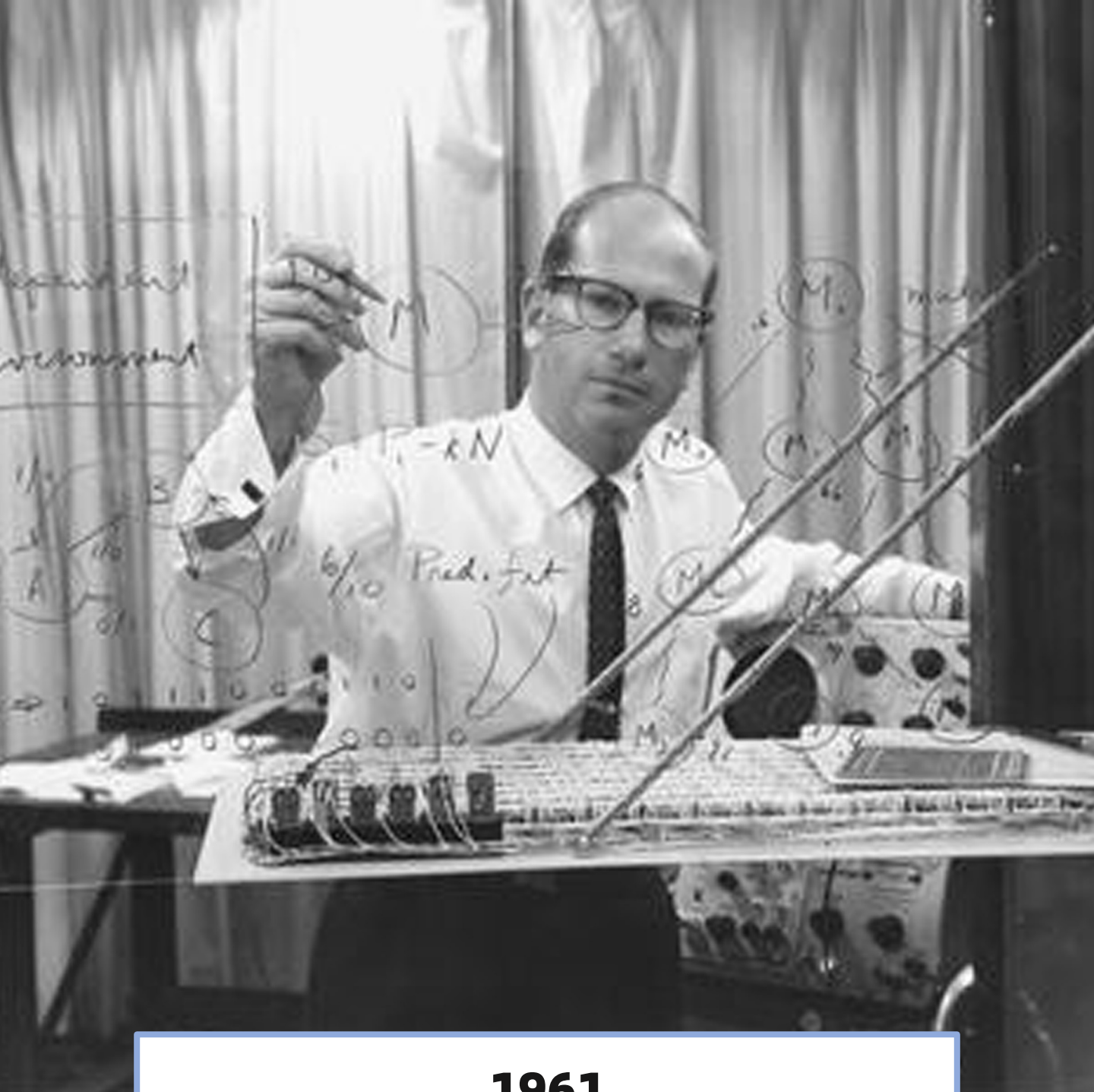


	Storm	Thunderstorm
Happy	$e_H(S)$	$e_H(T)$
Angry	$e_A(S)$	$e_A(T)$
Okay	$e_O(S)$	$e_O(T)$

1960

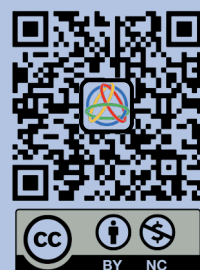
美国数学家伦纳德·E·鲍姆（Leonard Esau Baum）与1960年代，在IDA 通信研究部工作期间与Lloyd Welch一起开发了 Baum-Welch 算法。首次描述了描述了隐马尔可夫模型(HMM)。HMM 的首批应用之一是语音识别，始于 20 世纪 70 年代中期。





1961

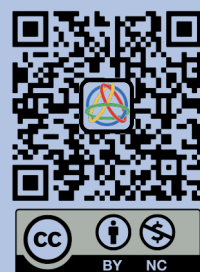
劳伦斯·小福格尔（Lawrence Jerome Fogel）在1961年，探索了时间序列预测的进化编程，被认为是人工智能进化计算的尝试。因此他也被称为主动降噪的发明者和进化编程之父。





1964

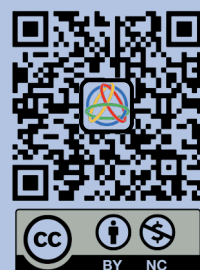
弗拉基米尔·瓦普尼克（Vladimir Vapnik）和阿列克谢·切沃年基斯（Alexey Chervonenkis）在1964年开发了Vapnik-Chervonenkis 理论，也是最初的SVM（支持向量机）算法。直到弗拉基米尔在1990年移居美国，他的算法才真正在机器学习社区发挥作用。





1967

1967年，马塞洛·佩利洛（Marcello Pelillo）发明“最近邻规则”。是基本模式识别的开端。该算法用于绘制路线，是最早用于解决旅行商寻找最有效路线问题的算法之一。使用它，销售人员输入一个选定的城市，并反复让程序访问最近的城市，直到访问完所有城市。

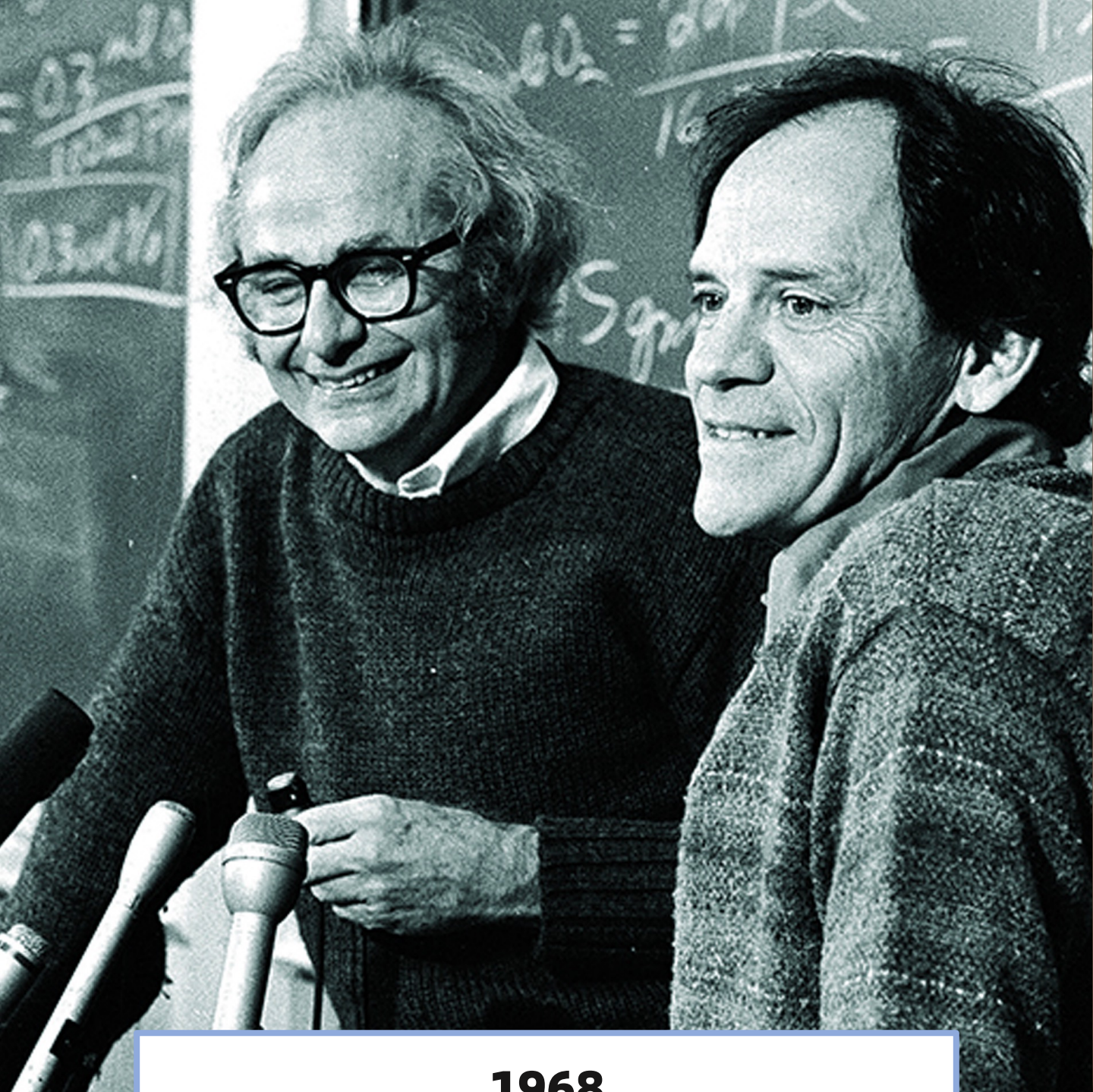




1968

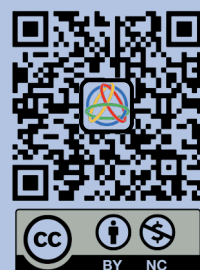
1968年，基辅控制论研究所的阿列克谢·伊瓦赫年科（Alexey G. Ivakhnenko）教授开发了数据处理组方法(GMDH)，这是一种归纳统计学习方法，因此他有时被称为“深度学习之父”。



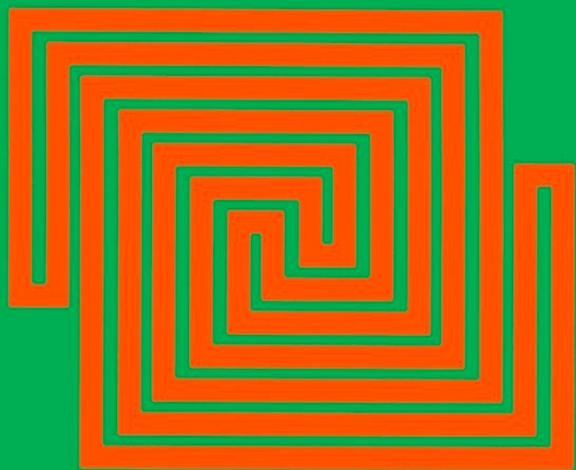


1968

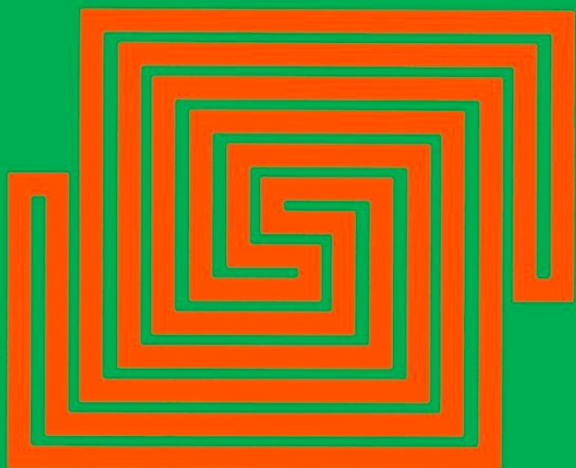
戴维·休贝尔（David H. Hubel）和托斯滕·威塞尔（Torsten Wiesel）在1968年发表了视觉皮层中的感受野的论文，介绍了两种视觉细胞：简单的单元格和复杂的细胞。这为后来的卷积神经网络算法奠定了理论基础。他们因此在1981年诺贝尔生理学或医学奖。



Expanded Edition



Perceptrons

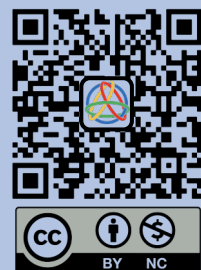


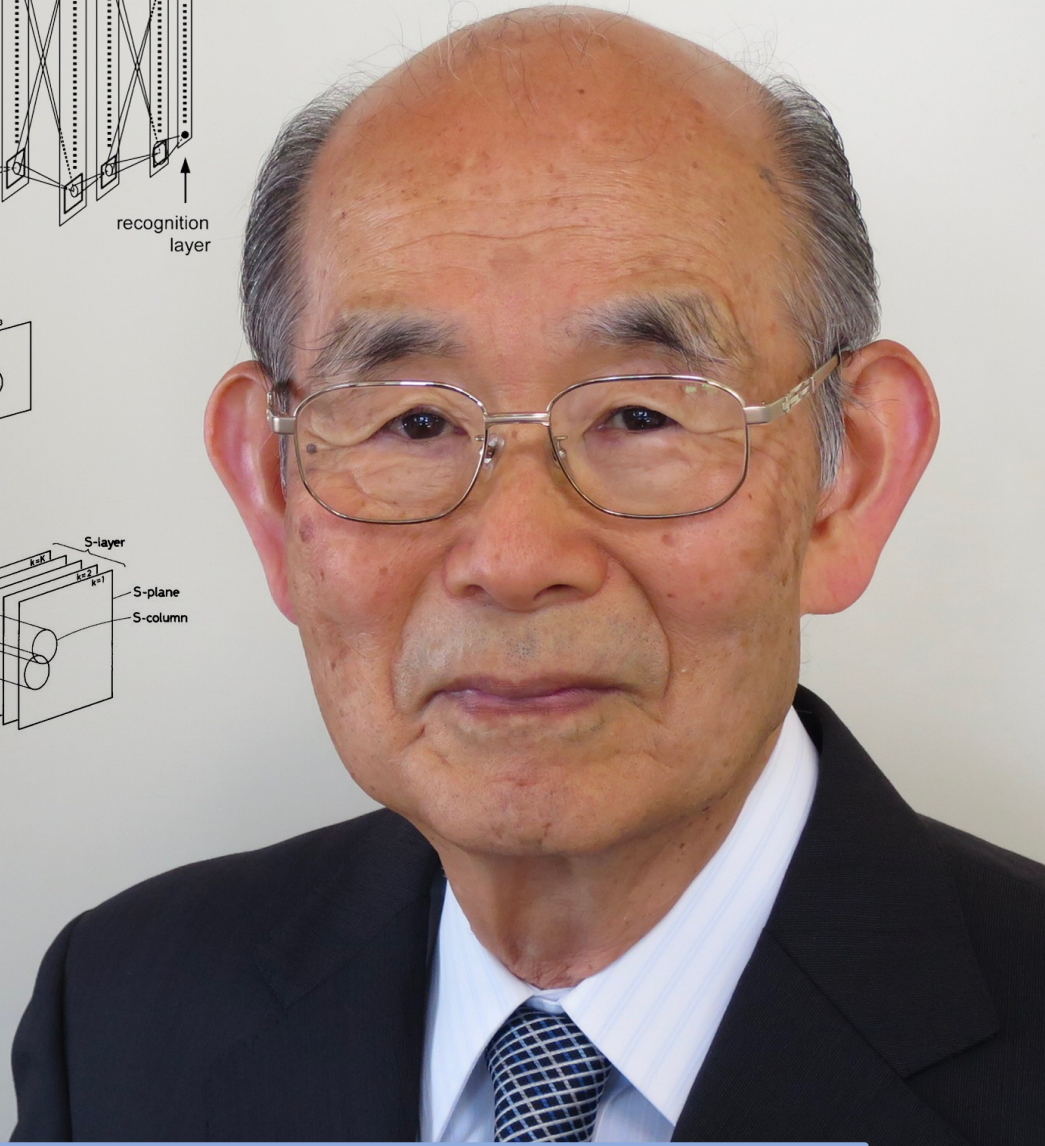
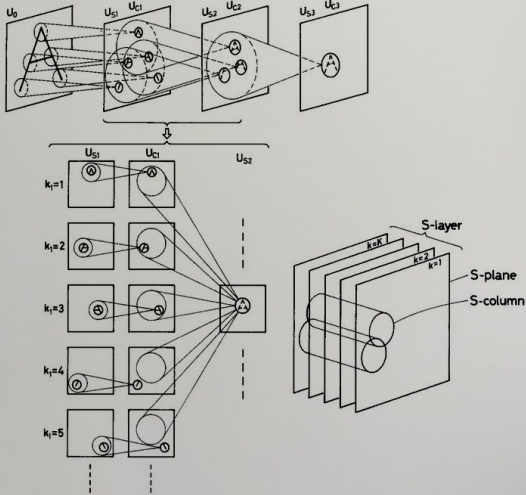
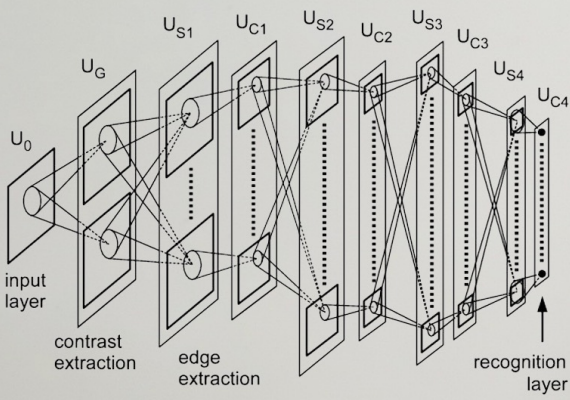
Marvin L. Minsky



1969

1969年，Marvin Minsky和Seymour Papert写了一本书《感知器》（Perceptrons）。这本书是献给心理学家 弗兰克·罗森布拉特的，他在1957年发表了第一个“感知器”模型。这本书争议很大。作者做出的悲观预测导致了AI研究方向的改变，并促成了1980年代长达十年的所谓的“AI 寒冬”。





1980

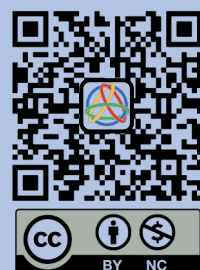
福島邦彦 (Kuniyuki Fukushima) 在1980年，发表了新认知机，这是最初的深度卷积神经网络(CNN) 架构。如今 CNN 架构通常通过反向传播进行训练。这种方法现在大量用于计算机视觉。





1982

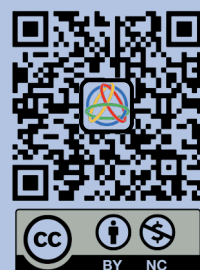
华中科技大学邓聚龙教授在1982年首创的一种系统科学理论(Grey Theory)，其中的灰色关联分析（GRA）是根据各因素变化曲线几何形状的相似程度，来判断因素之间关联程度的方法。这是人工智能领域一个著名的算法。





1984

伦敦帝国理工学院名誉教授伊戈尔·亚历山大（Igor Aleksander）在80年代启发设计了世界上第一个神经模式识别系统WISARD。以共同发明人 Bruce Wilkie、John Stonham 和 Igor Aleksander 的名字命名。他还设计了灵长类动物视觉系统、人类视觉语言系统、麻醉剂对意识的影响以及人工意识的神经模型。





Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California,
San Diego, La Jolla, California 92093, USA

† Department of Computer Science, Carnegie-Mellon University,
Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure¹.

There have been many attempts to design self-organizing neural networks. The aim is to find a powerful synaptic modification rule that will allow an arbitrarily connected neural network to develop an internal structure that is appropriate for a particular desired state the input uni

more difficult when we introduce hidden units whose actual or desired states are not specified by the task. (In perceptrons, there are 'feature analysers' between the input and output that are not true hidden units because their input connections are fixed by hand, so their states are completely determined by the input vector: they do not learn representations.) The learning procedure must decide under what circumstances the hidden units should be active in order to help achieve the desired input-output behaviour. This amounts to deciding what these units should represent. We demonstrate that a general purpose and relatively simple procedure is powerful enough to construct appropriate internal representations.

The simplest form of the learning procedure is for layered networks which have a layer of input units at the bottom; any number of intermediate layers; and a layer of output units at the top. Connections within a layer or from higher to lower layers are forbidden, but connections can skip intermediate layers. An input vector is presented to the network by setting the states of the input units. Then the states of the units in each layer are determined by applying equations (1) and (2) to the connections coming from lower layers. All units within a layer have their states set in parallel, but different layers have their states set sequentially, starting at the bottom and working upwards until the states of the output units are determined.

The total input, x_j , to unit j is a linear function of the outputs, y_i , of the units that are connected to j and of the weights, w_{ji} , on these connections

$$x_j = \sum y_i w_{ji} \quad (1)$$

put to each
in this extra

1985

戴维·鲁梅尔哈特 (David Rumelhart)、杰弗里·欣顿 (Geoffrey Hinton) 和罗纳德·J·威廉姆斯 (Ronald J. Williams) 在1985年写了一篇论文《反向传播算法》，引发了神经网络研究的热潮。该方法已广泛用于基础认知研究（如记忆、视觉识别）和实际应用。



PARALLEL DISTRIBUTED PROCESSING

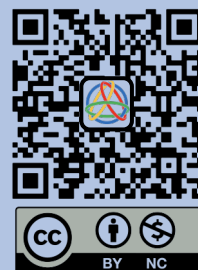
Explorations in the Microstructures of Cognition
Volume 2: Psychological and Biological Models



JAMES L. McCLELLAND, DAVID E. RUMELHART,
AND THE PDP RESEARCH GROUP

1985

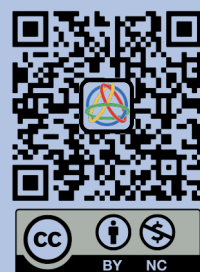
戴维·鲁梅尔哈特（David Rumelhart）和詹姆斯·麦克莱兰（James McClelland）一起发表了《并行分布式处理：认知微观结构探索》为计算机科学家提供了他们第一个可测试的神经处理模型，这是现在被视为认知科学领域的核心著作。描述了使用连接主义来模拟神经过程。





1986

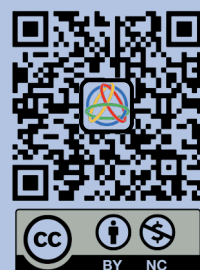
保罗·斯莫伦斯基（Paul Smolensky）在1986年以 Harmonium 的名字发明了受限玻尔兹曼机(RBM)。但在杰弗里·欣顿（Geoffrey Hinton）及其合作者于 2000 年为其发明了快速学习算法后声名鹊起。这个算法在深度学习网络中广泛运用。





1987

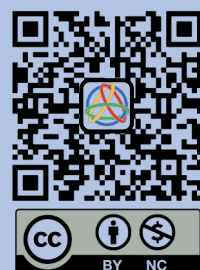
亚历克斯·韦贝尔 (Alex Waibel) 在1987年提出时间延迟神经网络 (TDNN)，它是第一个使用反向传播算法通过梯度下降训练的卷积神经网络 (CNN)。他是国际语音翻译研究联盟 C-STAR 的创始人之一。在 C-STAR，他的团队开发了 JANUS 语音翻译系统，这是第一个美国和欧洲语音翻译系统。





1989

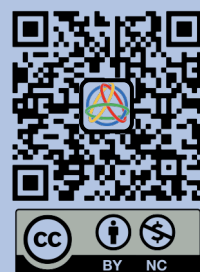
法国计算机科学家Yann André LeCun在1989年将标准反向传播算法应用于深度神经网络，使其成为卷积网络的奠基人，他也是DjVu图像压缩技术的主要创造者之一。





1991

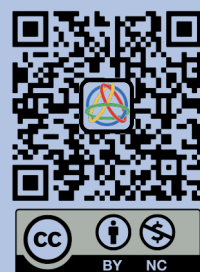
德国计算机科学家于尔根·施米德胡伯（Jürgen Schmidhuber）在1991年和学生们一起发布了一种称为长短期记忆(LSTM)的递归神经网络的论文。该论文分析并克服了著名的梯度消失问题。LSTM后续被广泛运用在语音识别领域。被称为“现代人工智能之父”，“成熟人工智能之父”，和“深度学习之父”。





1995

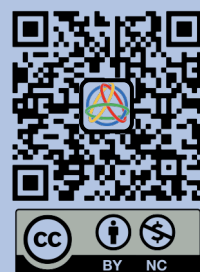
布伦丹·弗雷（Brendan Frey）在1995年发明了最早的深度学习方法之一：唤醒睡眠算法。20年后，他成立了一家公司叫Deep Genomics，这是第一家由人工智能发现的候选药物的公司。

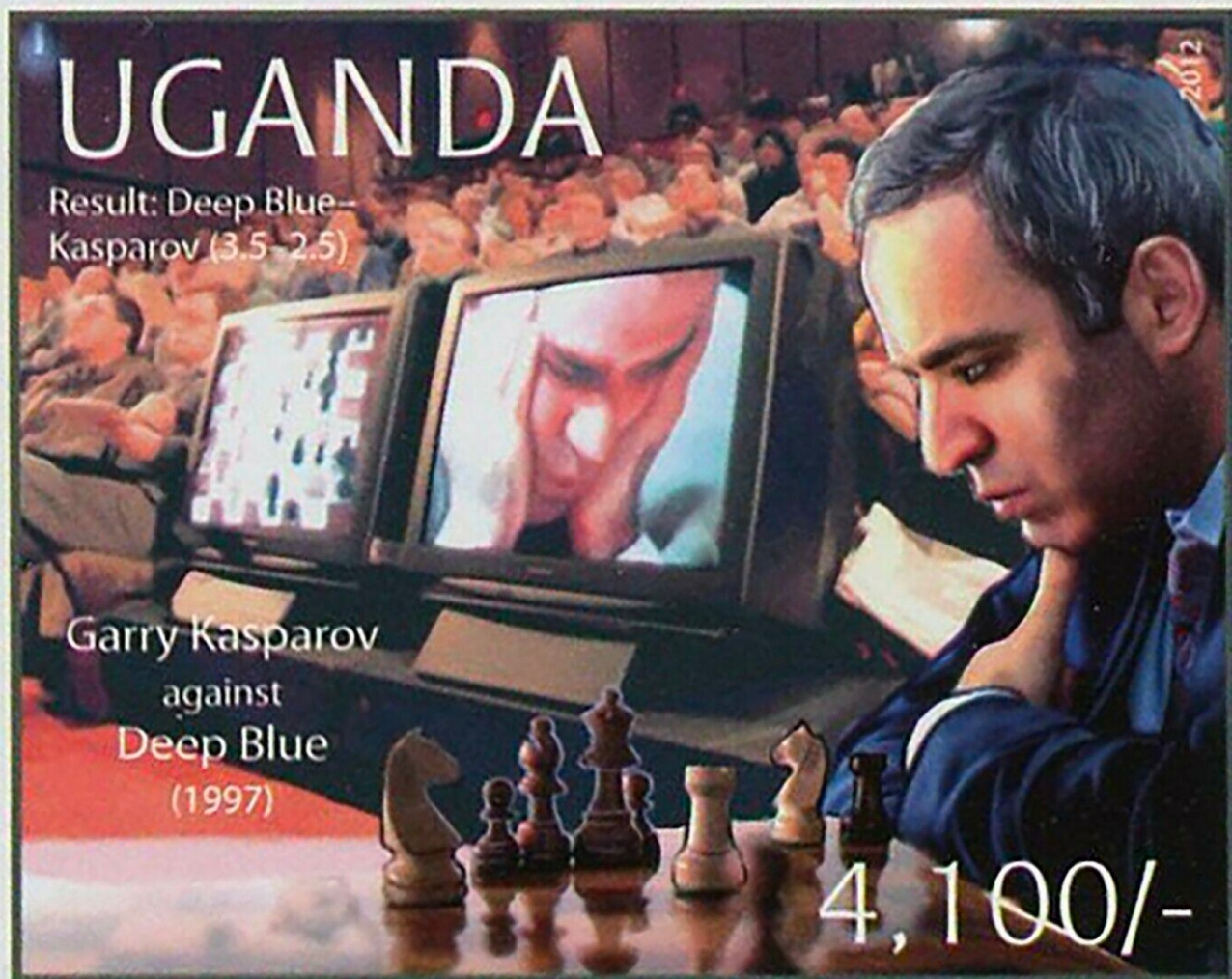




1995

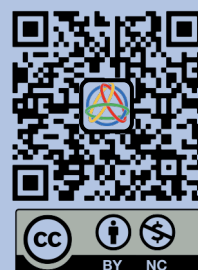
麻省理工学院教授罗莎琳德·皮卡德（Rosalind Picard）在1995年出版了《情感计算》一书。描述了情感在智力中的重要性，人类情感交流对人与人之间关系的重要作用，以及机器人和可穿戴计算机对情感识别的可能影响。开启了情感计算的领域。





1997

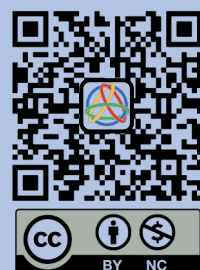
1997年，运行在IBM的Deep Blue深蓝国际象棋专家系统，击败了人类的国际象棋冠军，被认为是人工智能历史上的一个里程碑。





2000

DeepMind高级科学家马库斯·哈特（Marcus Hutter）在2000年提出了通用人工智能（AGI）的数学精确规范。他与于尔根·施米德胡伯（Jürgen Schmidhuber）和肖恩·莱格（Shane Legg）在2002年共同发表了人工智能的数学理论AIXI，该理论基于理想化智能代理和奖励激励强化学习。



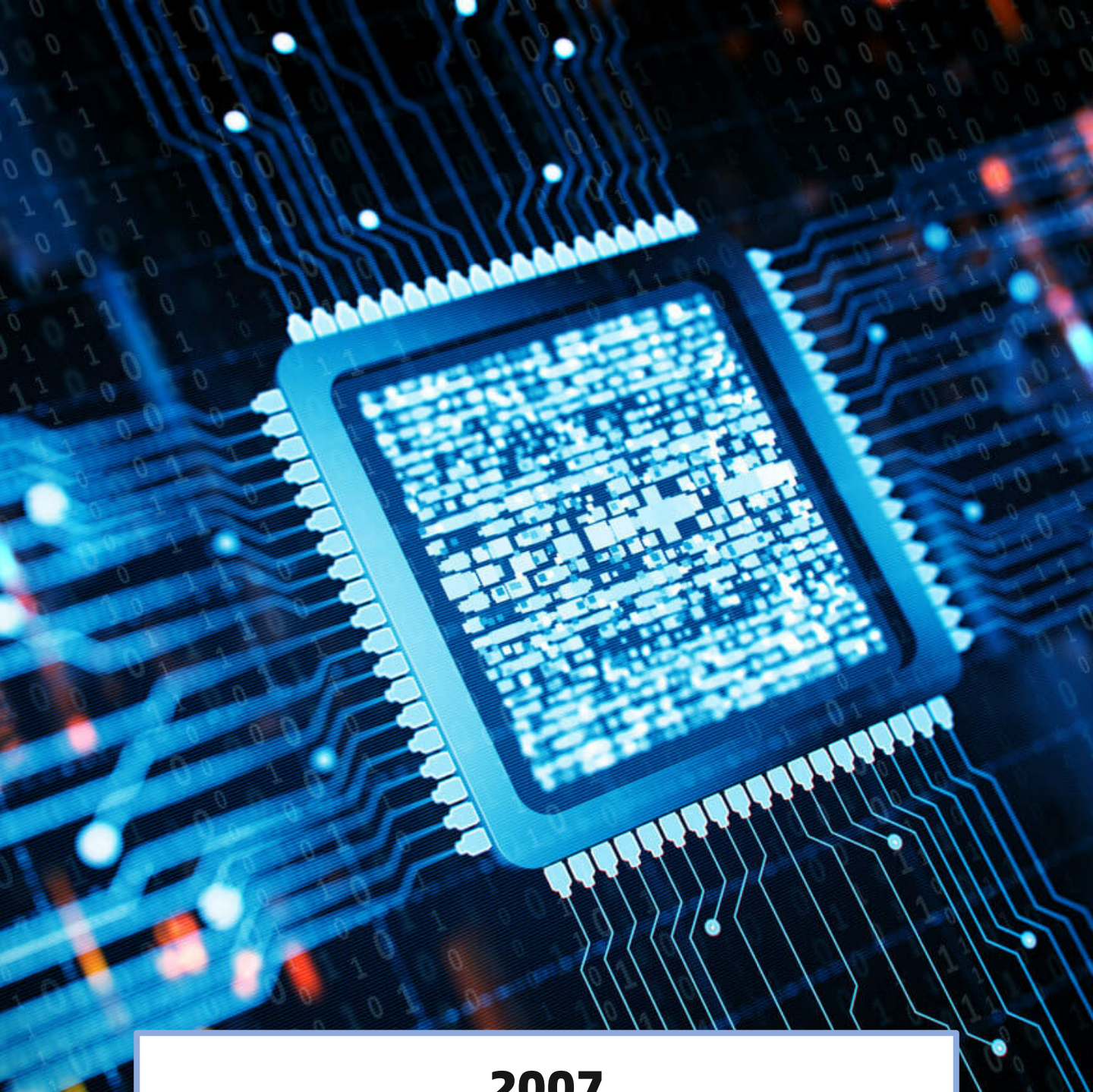
THE SINGULARITY IS NEAR



2005

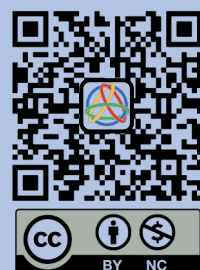
雷·库兹韦尔（Ray Kurzweil），在2005年写了一本非科幻小说《奇点临近》（The Singularity Is Near），第一次提出“奇点”，书中描述了他的加速回报法则，该法则预测计算机、遗传学、纳米技术、机器人技术和人工智能等技术将呈指数级增长。奇点也是机器智能与人类融合的点。并预测将奇点的日期定为 2045 年。





2007

2007年，随着 Nvidia GeForce 8 系列的推出，以及随后新的通用流处理单元，GPU 成为一种更通用的计算设备。GPU 上的通用计算开始真正进入机器学习领域。让标准神经网络在GPU实施比CPU上的等效实施快 20 倍。



Moral Machines

Teaching Robots Right from Wrong

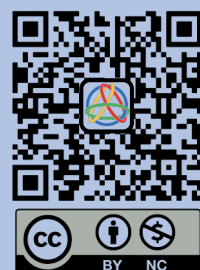


W e

2009

l e n

2009 年，牛津大学出版社出版了《道德机器，教育机器人明辨是非》，这是这是“第一本探讨构建人工道德主体所面临的挑战、深入探讨人类决策和伦理本质的书。书中引用了大约 450 个来源，其中约 100 个涉及机器伦理的主要问题。

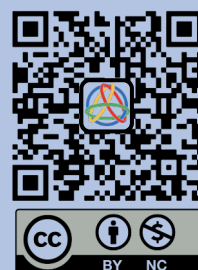




TensorFlow

2015

2015年，谷歌在Apache License 2.0下开源发布TensorFlow，这是一个用于机器学习和人工智能的免费开源 软件库。它可用于深度神经网络的训练和推理。这使得很多便宜的机器人也能实现面部识别和物体识别。





2017

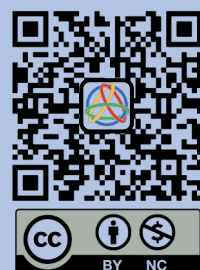
2017年，曾击败过李世石的AlphaGo又击败当时世界第一棋手柯洁，彻底完败人类。从2014年起开始下棋，到最后所向披靡，AlphaGo一共花了3年的时间。





2018

Tensor Processing Unit (TPU) 是谷歌为神经网络机器学习开发的AI加速器 专用集成电路 (ASIC) , 使用谷歌自家的TensorFlow软件。谷歌于 2018 年向第三方提供服务。某种程度刺激了机器学习的繁荣。

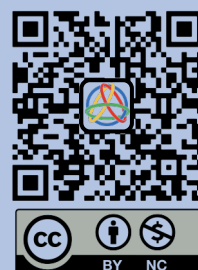


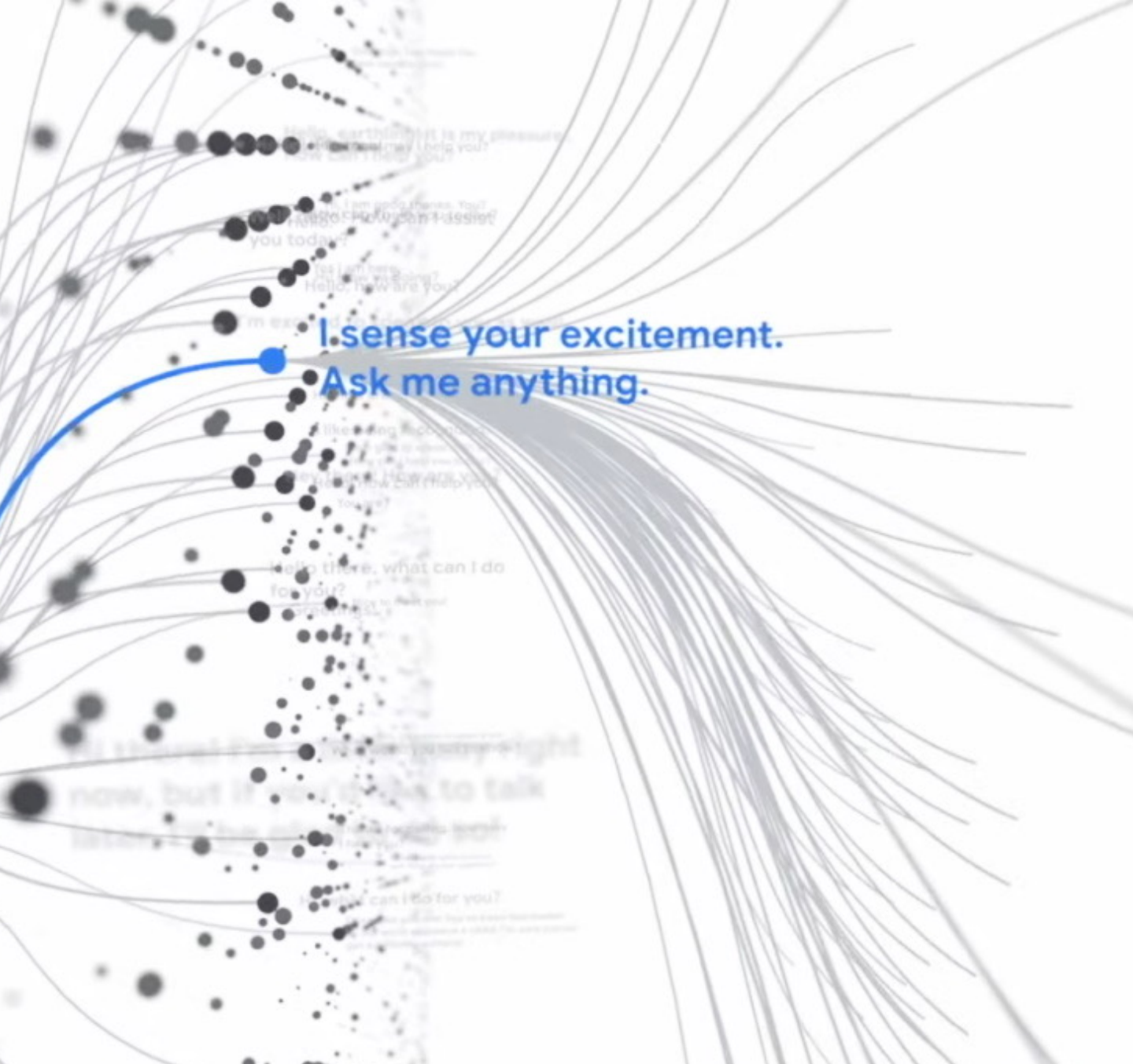


GTP-3

2020

总部位于旧金山的人工智能研究实验室OpenAI在2020年5月创建了第三代语言预测模型Generative Pre-trained Transformer 3 (GPT-3)。这是一种自回归语言模型，它使用深度学习来生成类似人类的文本。该架构是一个标准的转换器网络。具有前所未有的 2048 令牌 (Token) 长上下文和 1750 亿个参数。

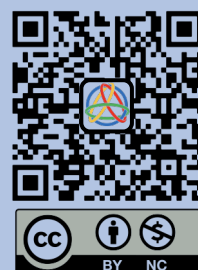




I sense your excitement.
Ask me anything.

2022

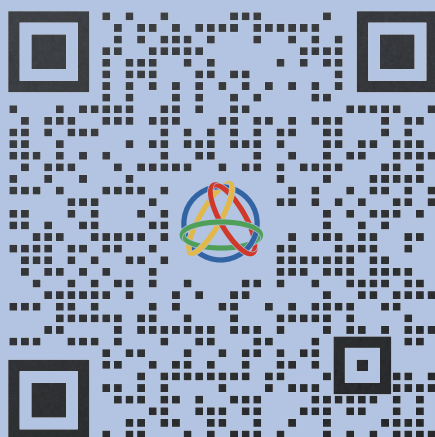
LaMDA是Google开发的一系列会话神经语言模型。2022年6月，当谷歌工程师Blake Lemoine 声称聊天机器人已经变得有知觉时，LaMDA获得了广泛关注。但谷歌否认了这个结论，随即解雇了Blake Lemoine，声称 Blake 违反了他们“保护产品信息”的政策。



关注我们，了解更多.....



扫码加入“共熵个大家庭”



深圳市共熵产业与标准创新服务中心
SHENZHEN GONGSHANG INDUSTRY AND STANDARD INNOVATION SERVICE CENTER