第 1 章 ChatGPT 发展简史

1000 多年前，人类就有制造智能机器的梦想。例如，传说中的木牛流马是由诸葛亮发明的一种运输工具，士兵驱使它在崎岖的栈道上运送军粮，而且“人不大劳，牛不饮食”。从描述上来看，这种机器应该是能够适应路况、自行移动的机器，类似现代的扫地机器人。然而，在当时的技术条件下，根本没有合适的工具来实现这个梦想。

在计算机出现之前，人们普遍认为人工智能的基本假设是可以将人类的思考过程机械化。直到电子计算机诞生，人们突然意识到，借助计算机也许可以实现建造智能机器的梦想，AI 就是基于计算机技术的一个新概念。

AI 已经发展了 70 多年，之前都是雷声大、雨点小，虽然各种各样的研究及讨论不断，但并没有激起太大的浪花。然而到了 2022 年 11 月 30 日，AI 的发展进入了一个新阶段——第一次被全球十几亿人热烈讨论，并在各个领域开始应用。因为，这一天 OpenAI 的 CEO 萨姆·奥尔特曼（Sam Altman）在推特上发布了 ChatGPT，一句话加上一个网站链接，任何人都可以注册账户，免费与 OpenAI 的新聊天机器人 ChatGPT 交谈。

24 小时内，大批人涌入网站，给 ChatGPT 提了各种要求。软件 CEO 兼工程师 Amjad Masad 要求它调试他的代码；美食博主兼网红 Gina Homolka 用它写了一份健康巧克力曲奇的食谱；Scale AI 的工程师 Riley Goodside 要求它为《宋飞正传》（Seinfeld）剧集编写剧本；Guy Parsons 是一名营销人员，他还经营着一家致力于 AI 艺术的在线画廊，他让它为他编写提示，以输入另一个 AI 系统 Midjourney，从文本描述创建图像；斯坦福大学医学院的皮肤科医生 Roxana Daneshjou 研究 AI 在医学上的应用，向它提出了医学问题，许多学生用它来做作业……

以前也出现过很多聊天机器人，但相形之下，比 ChatGPT 逊色不少。ChatGPT 可以进行长时间、流畅的对话，回答问题，并撰写人们要求的几乎任何类型的书面材料，包括商业计划、广告活动方案、诗歌、笑话、计算机代码和电影剧本。ChatGPT 可以在一秒内生成这些内容，用户无须等待，而且它生成的很多内容还不错。

它也会承认错误、质疑不正确的前提并拒绝不恰当的请求。

ChatGPT 看起来什么都懂，就像个百科全书。其流畅的回答、

丰富的知识，给了参与者极大的震撼。但它并不完美，也会产生让人啼笑皆非的错误、带来莫名的喜感。

ChatGPT 在发布后五天内，就拥有了超过 100 万名用户，这是脸书（现更名为 Meta）花了 10 个月才达到的里程碑。

2022 年 12 月 4 日，埃隆·马斯克（Elon Musk）发了一条推文，他说："ChatGPT 有一种让人毛骨悚然的厉害，我们离危险的强大人工智能已经不远了。"

仅仅 2 个月后，ChatGPT 注册用户突破 1 亿名。据网络流量数据网站 SimilarWeb 统计，ChatGPT 的全球访问量在 4 月份再创新高，达到 17.6 亿次。要知道，这个数字还没有包含大量使用 ChatGPT API 接入的和使用第三方应用的用户访问量。

ChatGPT 出现后，一夜之间，每个人都在谈论 AI 如何颠覆他们的工作和生活。

ChatGPT 是 AI 相关技术浪潮的一部分，这些技术被统称为"生成式人工智能"——其中包括热门的艺术生成器，如 Midjourney 和 Stable Diffusion。

在 ChatGPT 出现之前，大众对 OpenAI 了解很少，这家公司突然现身，引起了大家强烈的好奇心。它到底是什么来历？为什么能开发出如此强大的 AI 系统？

让我们从 AI 的起源说起。

1.1 AI 的概念和萌芽

1.1.1 什么是 AI？

在互联网上，AI（Artificial Intelligence，人工智能）有许多不同的定义。

维基百科说：AI 是指通过普通计算机程序来呈现人类智能的技术。

谷歌说：AI 是一组技术，使计算机能够执行各种高级功能，包括查看、理解和翻译口语及书面语言、分析数据、提出建议等各种能力。

百度百科说：AI 是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

Oracle 说：AI 是指可模仿人类智能来执行任务，并基于收集的信息对自身进行迭代式改进的系统和机器。

ChatGPT 说：AI 是一种机器智能，它可以执行通常需要人类智慧才能执行的任务，例如理解自然语言、解决问题、学习、决策和感知。AI 旨在通过模拟人类的智能和行为来创建智能机器和软件，研究领域包括机器学习、自然语言处理、计算机视觉、认知计算和智能机器人等。

为什么有如此多的定义呢？

因为 AI 是一个包括了多重内涵的概念，一个是指有 AI 的计算机（智能体），另一个是指 AI 科学和技术。分两点来解释：

第一，有 AI 的计算机可以通过模拟人类的智能和行为，完成通常需要人类智慧才能做的任务。

第二，AI 技术是让计算机成为一个持续学习提高的学生，不仅能从数据中学习和提取信息，并利用这些信息来做出正确的决策，还能感知环境，更好地理解周围的世界，更好地解决问题。

不过，AI 的概念依然存在很多争议。其中，最大的质疑是“智能”这个词。

顾名思义，AI 就是以人类智能为模板的，理解与模拟人类智能是 AI 实现的基本路径。

人类的智能首先表现在人类的身体能够感知和运动。通过眼睛、耳朵、鼻子、口和手脚这些感官，人类能够看到、听到、嗅到、品尝和触摸周围的事物，并通过手脚来操作和移动物体，与环境进行复杂的互动。机器要做到这些，需要具备识别模式和控制反馈的能力。

人类的智能更显现在人类复杂的心智活动中，例如理解语言、理解场景、规划行动、智能搜索信息、学习总结知识、推理决策等，这种高级智能活动由人的大脑完成。如果 AI 只停留在感知和运动阶段，那么机器只能达到普通动物的智力水平，而人的心智能力才是让人类成为万物之灵的关键。

1.1.2　AI 的分类

2016 年 AlphaGo 打败了人类围棋冠军，按理说比人类还强的机器棋手具有了相当强大的推理和思考能力，那么它算拥有智能了吗？

当然算。因为它模拟出来了人类的智能，具有强大的计算和搜索技术，可以通过分析和模仿大量数据来学习并识别围棋中的模式和策略，以做出最佳的棋局决策。但是 AlphaGo 并没有像人类一样的推理和思考能力，无法像人类那样通过推理进行逻辑推断，所以 AlphaGo 也无法将下棋的高超技能迁移到其他领域。换句话说，下棋赢过所有人的 AlphaGo 在写文章方面还不如小学生。这种 AI，被称作弱 AI。弱 AI（Weak AI）并不是说 AI 本身的能力很弱，而是指它只能解决少量的问题。实际上，目前所有的 AI 都只能完成单个领域的工作，都叫作弱 AI，也叫窄 AI（Narrow AI）。

生活中，弱 AI 的应用随处可见，场景也越来越广泛，为人们的生活、工作和社会带来了巨大的改变和进步：

拍照时，通过美颜相机里的一键美颜功能修改照片；

睡觉时，通过语音发出指令，控制灯光、空调等，方便而快捷；

购物时，随时通过智能客服直接获取帮助，避免长时间等待；

旅游时，通过人脸识别设备确定身份，快速通过安检或自助入住酒店。

而能写出大量文章、貌似无所不知的 ChatGPT，也是弱 AI。它在推理上常常会犯一些基本错误，且其逻辑推理能力很差。

虽然现在人们生活中用到了各种 AI 产品，不断有 AI 相关的新闻和事件，但迄今为止还没有任何 AI 通过图灵测试。

实际上，AI 的发展过程可以类比为一个孩子的成长过程。刚开始，它只是一个没有多少知识的婴儿，需要人类的指导和培育；接着，它逐渐成长为一个会爬会走的孩子，可以自主探索和学习；然后，它成长为一个可以独立思考和解决问题的少年，开始展现出自己的才华和潜力；最终，它成为一个成熟的、有着强大智慧和创造力的成年人，为人类带来巨大的改变和进步。

现在，AI 正处于少年期，还远远没有到成熟期，但它已经在很多领域展现出了强大的能力和潜力，如语音识别、图像识别、自然语言处理、自动驾驶等。

人们预测，随着 AI 的继续发展，会出现强 AI（Artificial General Intelligence，AGI），也叫通用人工智能。这是指跟人类的能力相似的 AI，它是能够像人类一样思考、理解、推理和学习等的系统。比起弱 AI，强 AI 更全面，它能通过观察、感知和处理外部世界的信息，独立、自主地做出决策和行动。例如，在电影《星球大战》中的智能机器人 BB-8（如图 1-1 所示），它可以在没有人类指示和操控的情况下独立思考和解决问题，还可以自我学习和进化，最终实现自主决策和行动。

图 1-1　BB-8 机器人（来源：电影《星球大战：原力觉醒》）

目前，ChatGPT 已经具备了不少了强 AI 的特征，有希望发展成为强 AI。

假设计算机程序通过不断发展，可以比世界上最聪明、最有天赋的人类还聪明，那么，由此产生的 AI 系统就可以被称为超 AI（Artificial Super Intelligence，ASI），也就是超越人类的 AI。显然，现阶段这只是一种存在于科幻电影中的想象场景。未来学家和科幻作者喜欢用“奇点”来表示超 AI 到来的那个神秘时刻，但没有人知道奇点会不会到来，会在何时到来。

与信息技术产业的其他分支（如计算机、通信、互联网、智能手机等）所取得的科技成果比较，AI 领域所取得的成果还远远不够，未能达到最初的设想。

虽然在 AI 的发展过程中人才辈出，但在很长一段时间内，研究者们一直像在迷宫中探索，没有人知道清晰的路径。有些看似走得通的路，走着走着就走进了死胡同；更有趣的是，一些开始认为行不通的路，过了一段时间，发现居然又能走通了。原因非常复杂，影响 AI 的关键因素多，路径多，导致 AI 的发展几经波折，起起落落。

人们设想中的“AI”能模仿人类，具有高级的“推理”“思考”或“认知”能力。然而，经过 70 多年的探索，高层次的推理和思考仍然难以捉摸。

为什么 AI 领域的进展会不如预期，甚至在几个特定的时间段内完全陷入“泥潭”？让机器拥有智能这么难吗？人类精英们都做出了哪些努力和探索？目前的 AI 到达了什么阶段？

下面，我们从 AI 之父图灵开始，回顾 AI 的发展历史，在本书 1.5 小节中可再进一步了解 ChatGPT 的发展过程。

1.1.3　图灵测试

让计算机像人一样说话，是在计算机发明之前的梦想。

1942 年，科幻小说作家艾萨克·阿西莫夫（Isaac Asimov）在短篇小说《转圈圈》（*Run around*）中创造了机器人（Robot）这个词，小说中的机器人吉斯卡拥有比普通机器人更高的智能和自主性，不仅可以用人类的语言还能用眼神、手势和肢体语言等方式进行交流，这使他更加接近人类，也更容易被人类接受和信任。小说提出了“机器人三大法则”：第一条，机器人不得伤害人类或看到人类受到伤害而袖手旁观；第二条，机器人必须服从人类的命令，除非这条命令与第一条相矛盾；第三条，机器人必须保护自己，除非这种保护与以上两条相矛盾。后来这三大法则被许多小说、电视剧等广泛借用。

1950 年 10 月，英国数学家、逻辑学家艾伦·图灵（Alan Turing，见图 1-2）发表了一篇划时代的论文《计算机与智能》（*Computing Machinery and Intelligence*），提出了一个有趣的问题：假如有一台宣称自己会“思考”的计算机，人们该如何辨别计算机是否真会思考呢？

图 1-2　1951 年的艾伦·图灵（来源：nytimes.com）

为此，他设计了一个实验：如果一台机器与人展开对话，测试者分不清幕后的对话者是人还是机器，那么可以说这台机器具有智能。让机器模仿人就是图灵所说的“模仿游戏”，后来这个设想也被人们称为“图灵测试”。

随后，图灵又发表了论文《智能机器，被视为异端的理论》（*Intelligent Machinery, A Heretical Theory*）。因为图灵的这两篇论文探讨了机器具有智能的可能性，并对其后的机器智能发展做了大胆预测，所以他被称为“AI 之父”。从那时开始，70 多年来，人类一直试图解决这个问题，希望能研发出可以通过图灵测试的 AI。

这个梦想也驱使许多人持续探索。早在 1966 年，MIT 的教授约瑟夫·维森班（Joseph Weizenbaum）就开发了第一个聊天程序 ELIZA，之后的 50 多年，陆续出现了更先进的 AI 机器人，如微软小冰、苹果 Siri、谷歌助手、小度音箱等（聊天机器人发展时间线如图 1-3 所示）。

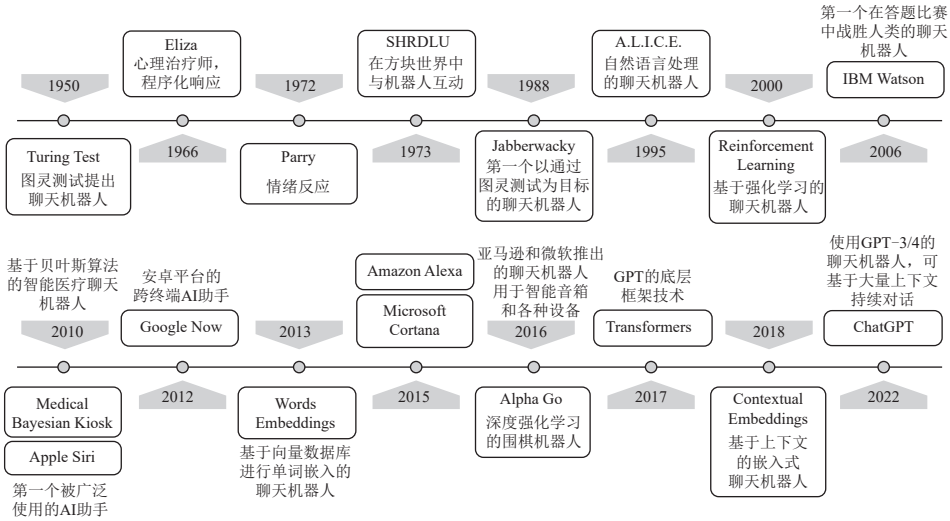


图 1-3　聊天机器人发展时间线（来源：hellofuture.orange.com）

但直到现在，计算机还不能像真人一样聊天。

2014 年 6 月 8 日，英国雷丁大学在著名的伦敦皇家学会举办了一场图灵测试。在当天的测试中，一组人类裁判以键盘输入的形式与电脑“对话”。如果裁判认定电脑为人的比例超过 30%，则电脑通过测试。5 个参赛电脑程序之一的尤金·古兹曼（Eugene Gootsman）成功“伪装”成一名 13 岁男孩，在一次为时 5 分钟的文字交流中，回答了裁判输入的所有问题，其中 33% 的裁判认为与他们对话的是人而非机器（如图 1-4 所示）。

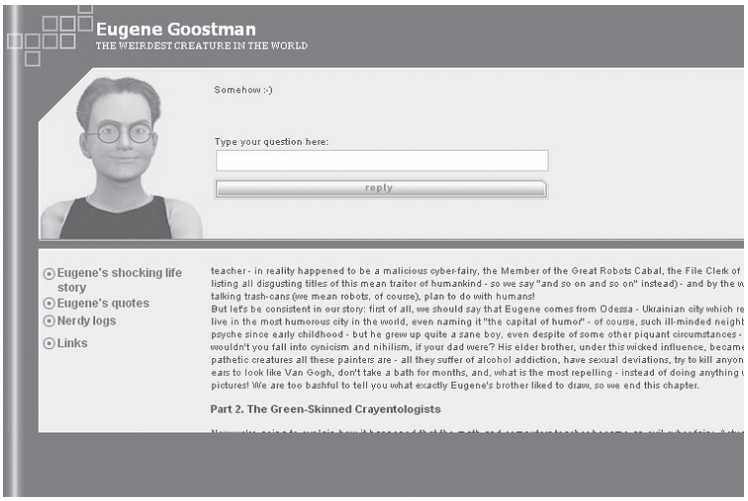


图 1-4　与尤金·古兹曼聊天的界面

有人认为，这个程序通过了图灵测试，成为有史以来第一个具有人类思考能力的 AI。

也有人质疑，这个测试的提问时间短，裁判少，严格来说，不能算通过了图灵测试。

大家的共识是，到目前为止，还没有任何 AI 通过图灵测试，而最接近通过图灵测试的就是 ChatGPT。

许多人认为，对 ChatGPT 这个每天都在跟人对话中学习的 AI 来说，通过图灵测试应该只是时间问题。

1.1.4　AI 的萌芽

1936 年 9 月，图灵应邀到美国普林斯顿大学高级研究院学习，两年后获得博士学位。也许图灵自己都不会想到，十几年后，普林斯顿大学成为 AI 大牛们最重要的启蒙地。

1949 年 9 月，约翰 • 麦卡锡（John McCarthy）来到普林斯顿大学研究数学，也是两年获得博士学位。麦卡锡希望机器能够像大脑一样做到推理知识，普林斯顿大学高级研究所的约翰 • 冯 • 诺伊曼（John von Neumannn，现代计算机之父）鼓励麦卡锡写下他的想法。1952 年夏天，麦卡锡去贝尔实验室工作，遇到克劳德 • 香农（Claude Shannon），他们合作撰写了大量关于自动机的论文。

正是在普林斯顿大学，麦卡锡第一次遇到了攻读数学博士的马文 • 明斯基（Marvin Lee Minsky），他们都对机械智能很感兴趣。

1951 年，马文 • 明斯基和迪安 • 埃德蒙兹（Dean Edmonds）设计制造了第一台随机连接神经网络学习机 SNARC（Stochastic Neural Analog Reinforcement Calculator，如图 1-5 所示），它有一个很酷的别名：“谜题解决者”（Maze Solver）。机器由 400 个真空管制成，可以在一个奖励系统的帮助下完成穿越迷宫的游戏。1954 年明斯基在普林斯顿的博士论文题目是《神经—模拟强化系统的理论及其在大脑模型问题上的应用》（*Theory of Neural-Analog Reinforcement Systems and its Application to the Brain-Model Problem*），是一篇关于神经网络的论文。

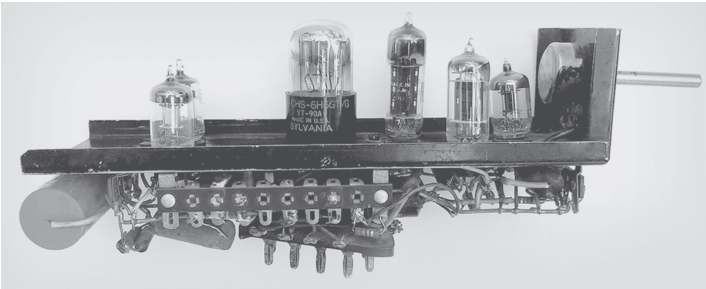


图 1-5　SNARC（来源：the-scientist.com）

1.1.5　AI 学科成立

当时类似香农、麦卡锡、明斯基这样 AI 方向的研究者们都在不同的地方各自研究，零零散散，比较孤立，能否把大家聚在一起，创立一门新学科呢？

明斯基从普林斯顿大学毕业后，去哈佛大学担任数学与神经学初级研究员，麦卡锡在达特

茅斯大学做数学助理教授，两人一拍即合，决定召集一个会议。

于是，AI 学科诞生的里程碑——达特茅斯会议，就由这两个年轻人发起了，会议计划书是麦卡锡写的，总结则由明斯基完成。

麦卡锡（如图 1-6 所示）在为研讨会写提案时创造了“AI”一词，他在会议的提案中说，研讨会将探索这样的假设：“（人类）学习的每一个方面或智能的任何其他特征原则上都可以被精确描述，以至于可以用机器来模拟它。”

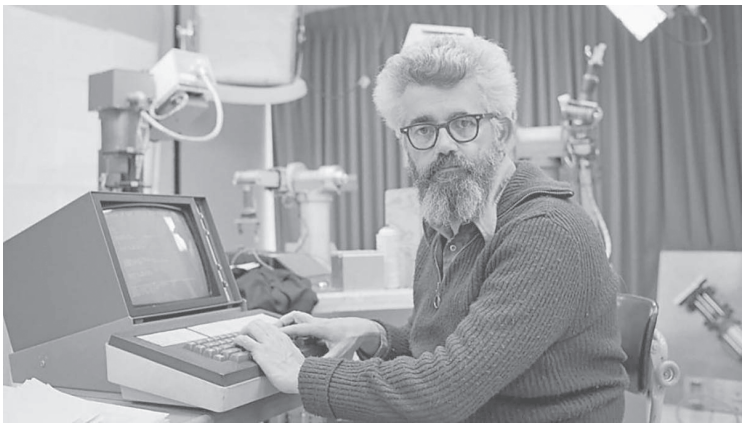


图 1-6　约翰・麦卡锡（来源：analyticsdrift.com）

会议组织起来并不容易，虽然场地由达特茅斯大学免费提供，但没有经费，而且两个组织者的地位不高，一个初级研究员，一个助理教授，在学术界都只能算小人物。这两个年轻人组织会议，估计许多“大牛”不会来，所以他们又找到了两位“大神”担任发起人：一个发起人是信息论的创始人克劳德・香农，他当时已经是贝尔实验室的大佬；另一个是 IBM 的资深专家纳撒尼尔・罗切斯特（Nathaniel Rochester），他是世界上第一台大规模生产的科学用计算机 IBM701 的首席设计师，编写了世界上第一个汇编程序。

由于会议时间长达两个月，为了让大家能安心开会，麦卡锡觉得应该给些生活费，于是他向洛克菲勒基金会申请赞助，他提出的预算是 13500 美元，但只获批了 7500 美元，由此可见当时这个会议并没有受到关注。

1956 年 7 月，一批数学家和计算机科学家（如图 1-7 所示）来到达特茅斯学院数学系所在大楼的顶层。在大约八周的时间里，讨论用机器来模仿人类学习以及其他方面的智能等问题，之后又陆陆续续来了一些人阶段性参与。

正是在这次会议上，AI 的概念第一次被提出。从事后看来，这次会议是金庸小说中“华山论剑”级别的。因为在主要的十位参会者中，有两位成名已久的人物——香农和纳撒尼尔，其余八人中，有一位机器学习之父亚瑟・塞缪尔（Arthur Samuel）以及四位图灵奖得主：赫伯特・西蒙（Herbert A. Simon）、艾伦・纽厄尔（Allen Newell）、马文・明斯基 、约翰・麦卡锡。赫伯特・西蒙给自己取了一个中文名叫司马贺，下面我们都以司马贺代称他。

图 1-7　达特茅斯会议的部分参会者（来源：roboticsbiz.com）

四人中，纽厄尔、明斯基和麦卡锡是同龄人（1927 年出生），当时都才 29 岁，司马贺 40 岁，算大龄青年，他们年轻气盛，野心十足，不肯妥协。

虽然麦卡锡召集大会的时候用了 AI 这个名字作为新学科的名字，但纽厄尔和司马贺却不认可，他们主张用“复杂信息处理”（Complex Information Processing）这个词作为学科研究方向的专业名词，他们的理由是“人工”（Artificial）一词并不能体现对机器实现智能研究的初衷，这个学科明明是研究计算机怎样才能自动处理信息的，加上“人工”不就变味了？这理由听起来也有道理。

但换个角度看问题，Artificial 在英文中也有“人造的”“模拟的”等意思，也是很贴切的，反而译成中文显得有点怪。不过，最终还是就“AI”这个名字达成了共识，这也是这次大会的一个重要成果。

到今天，AI 一词已被大众普遍接受了，还有人半开玩笑地给出了新的解释，AI 是有多少“人工”才有多少“智能”。确实，李飞飞的 ImageNet 图库用了大量的人工来做标注，才能形成给 AI 的学习语料，而 ChatGPT 用于引导模型方向的人工标注量也是非常惊人，还有 Scale.ai 这样专门做 AI 数据标注的公司，估值高达 73 亿美元。

早期 AI 领域并没有大家公认的学术领袖，更没有明确的研究路径，所有人只能在摸索中前进。逐渐分成了三个主要流派：符号主义、连接主义、行为主义，当然这些方向的概念也都是后面总结的。

实际上，这三个流派是对人类三种学习形式的模仿。

（1）符号主义流派借鉴了人类学习语言和数学的方式，比如借助一些文字、数学中的符号、图示以及概念等，帮助人们理解和思考各种问题，从而增强认知能力。然后，通过掌握一些规则，就能形成智能，比如用规定的字母和符号，按照一定的规则就能写出一篇文章。因此符号主义强调思维和认知过程中的符号表示和处理，这个 AI 的实现路径也是图灵的想法。

（2）连接主义流派借鉴了人的大脑生理结构，因为那时候人们已经认识到人类的大脑主要结构就是神经元构成的神经网络，多个神经元之间相互连接，可以获取并存储知识，完成思考过程。脑部和神经元的连接方式和权重的调整对人类学习和思考起着重要作用，假如能创建出人工神经元，再连接在一起，经过一段时间的学习，就能形成智能，因此连接主义在 AI 领域中主要强调神经网络连接的不同之处和深度学习。

（3）行为主义流派借鉴了人的学习受到反馈的影响，比如在学生学习过程中，老师对答题的评判、家长的一些惩罚奖励和外部环境等都能影响孩子的学习成绩和掌握知识的程度。也就是说，人们通过学习中的反馈来调整自己的行为和习惯，从而积累知识和经验，形成智能。因此，行为主义在 AI 领域中主要强调学习者的行为和环境的响应，并利用相关技术改变个体行为习惯。

大致来说，符号主义更多地模仿了人对语言等知识的学习策略，连接主义模仿了人的大脑生理结构，而行为主义更多地模仿了人受到外部激励后的反应。

现在我们知道，这些都对、都有用，而且可以混着用。但在达特茅斯会议时，符号主义占了压倒性优势，四位图灵奖得主全都是符号主义的代表。可惜的是，符号主义在发展了 30 多年后，碰到了瓶颈，现在已经不是 AI 的主流了。这是怎么回事呢？

1.2 符号主义流派的探索

符号主义（Symbolicism）又叫逻辑主义，其起源可以追溯到微积分的发明人莱布尼茨。他曾提出一个问题：“我们每天看到的世界，是通过具体化的、非结构化的形式（如视觉、声音、文字等）呈现的，这些具体却散漫粗疏的内容是否可以使用符号来精确定义？人类理解世界的过程，是否可以在这个基础上精确描述和计算？”

符号主义流派的学者们的答案是“可以”。

符号主义流派的核心思想可以概括为五个字：认知即计算。也就说，计算机可以通过符号模拟人类的认知程度来实现对人类智能的模拟。比如，人学习新词汇时，需要理解这些符号（即单词）的含义以及它们在文本中的作用。对计算机来说，单词、句子、数字、标点符号和语法规则等都是符号，先明确符号的含义，通过程序让计算机理解这些符号后，然后进行计算和推理，思考过程就是符号的操作过程。这种方法非常适合用来处理一些逻辑上的问题，比如证明定理、语义表达等。

1.2.1 符号主义的“推理期”（1956—1965 年）

在达特茅斯会议上，纽厄尔和司马贺（如图 1-8 所示）发布了“逻辑理论家（Logic Theorist）”，给所有人留下了深刻的印象。明斯基说那是“第一个可工作的人工智能程序”，因为这个程序可以证明怀特海和罗素《数学原理》中第二章 52 个定理中的 38 个定理。

从现在的视角看，逻辑理论家采用的算法和计算机自动定理证明这个学科分支的主流算法思想（如归结原理、DPLL 算法）并没有很直接的关系，甚至可以说只是对形式化的前提进行简单粗暴的搜索而已，技术含量并不高，但在当时这已经让人非常兴奋了。

图 1-8　艾伦・纽厄尔（左），司马贺（右）（来源：timetoast.com）

纽厄尔和司马贺借鉴了心理学的研究方法，把人脑看成一个实现信息处理目的的物理符号系统来表现智能行为，提出了“物理符号系统假说”（Physical Symbol System Hypothesis，PSSH）。1975 年，两人共同获得了图灵奖，以表彰他们在 AI、人类识别心理和表处理方面的贡献。

而另外一位参会者，来自 IBM 的亚瑟・塞缪尔也拿出了自己的成果。1952 年他为 IBM 701 的原型编写了 AI 跳棋程序，1955 年他增加了一些功能，使该游戏能够从经验中学习，即让程序记住以前游戏中的好走法，然后在下棋时快速搜索。

为什么科学家喜欢研究下棋游戏呢？

一方面是因为棋类游戏自古以来就被认为是人类智力活动的象征，模拟人类活动的 AI 自然要以此为目标，成功达到人类甚至高于人类水平，这就可以吸引更多人关注并投身于 AI 的研究和应用中。

另一方面，棋类也很适合作为新的 AI 算法的标杆，因为棋类游戏的规则简洁明了，输赢都在盘面上，适合计算机来求解。理论上只要在计算能力和算法上有新的突破，任何新的棋类游戏都可能得到攻克。

麦卡锡在那个时期的主要研究方向也是计算机下棋，为了减少计算机需要考虑的棋步，麦卡锡提出了著名的 α-β 剪枝算法的雏形。随后，卡内基梅隆大学的纽厄尔、司马贺等很快在实战中实现了这一技术。

α-β 剪枝算法的名字来源于它的主要思想——剪枝。在搜索树的任意一位置，如果计算机发现其中某些分支不会对最终的结果产生影响，即已经找到了更好的解决方案或者已经发现了相反的答案，那么这些分支就可以被剪掉，从而减少搜索时间，因此这种算法被称为剪枝算法。在 α-β 剪枝中，α 和 β 是两个重要的变量，其值在剪枝过程中发挥了关键作用。

由于采用了 α-β 剪枝算法，塞缪尔的跳棋程序曾在 1962 年赢得了一场对前康涅狄格州跳棋

冠军罗伯特·尼尔利的比赛，这在当时可以说是非常辉煌的成绩。

在相当长的一段时间内，α-β 剪枝算法成了计算机下棋的主要算法框架，1997 年战胜国际象棋大师卡斯帕罗夫的 IBM 深蓝也采用了 α-β 剪枝算法。

在 AI 早期，通过算法优化，让机器智能胜过人类，确实是让人惊喜的。

在达特茅斯会议之后的十多年里，AI 蓬勃发展，被广泛应用于数学和自然语言领域，用来解决代数、几何和英语问题，这个时期称为符号主义的“推理期”。因为研究者们重点解决的问题是利用现有的知识去做复杂的推理、规划、逻辑运算和判断，“推理”和“搜索”成为 AI 的思维研究方向。推理就是通过人类的经验，基于逻辑或事实归纳出来一些规则，然后通过编写程序来让计算机完成一个任务。搜索则属于“暴力计算”，具体来说，计算机可以通过在解决问题的途径上设置节点，并对各个节点进行前后逻辑的持续分析，在不厌其烦的试错下，最后根据指示找到正确的目标。

1958 年，明斯基从哈佛大学转到 MIT（麻省理工学院），同时麦卡锡也由达特茅斯学院来到 MIT 与他会合，他们在这里共同创建了世界上第一个 AI 实验室。同年，麦卡锡发明了 Lisp 语言，这是 AI 界第一个最广泛流行的语言，至今仍在广泛应用，它与 1973 年出现的逻辑式语言 PROLOG 并称为 AI 的两大语言。

两年后，麦卡锡第一次提出将计算机批处理方式改造成分时方式，这使得计算机能同时允许数十甚至上百用户使用，极大地推动了接下来的 AI 研究。这种分时策略也促进了各类计算机的发展，比尔·盖茨就是通过一个终端设备接入小型计算机，从而掌握了编程技术的。

由于这些成就，麦卡锡获得了 1971 年的图灵奖。

1961 年，第一台工业机器人 Unimate（如图 1-9 所示）在新泽西州通用汽车的组装线上投入使用，它看起来就是一个机械臂，能够运输压铸件并将其焊接到位。

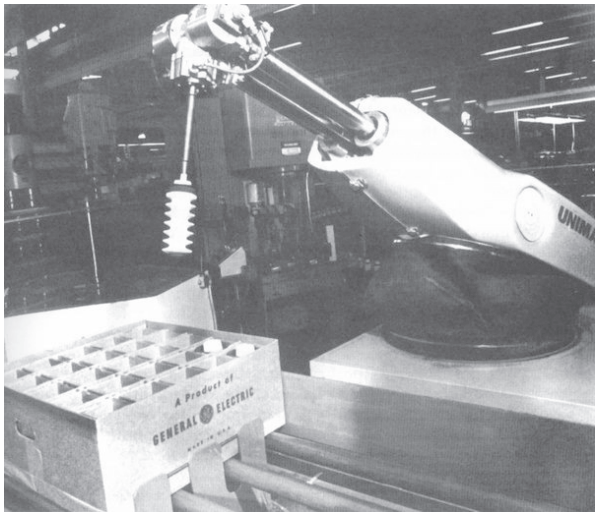


图 1-9　工业机器人 Unimate（来源：computerhistory.org）

1966 年，MIT 的教授约瑟夫·维森班（Joseph Weizenbaum）开发了第一个聊天程序 ELIZA，它可以根据任何主题进行英文对话。

ELIZA 的原理非常简单，在一个有限的话题库里，用关键字映射的方式，根据问话，找到自己的回答。

比如，在对话小程序通过谈话帮助病人完成心理恢复这个场景，当用户说“你好”时，ELIZA 就说：“我很好。跟我说说你的情况。”此外，ELIZA 会用“为什么？”“请详细解释一下”之类引导性的句子来让整个对话不停地持续下去。同时 ELIZA 还有一个非常聪明的技巧，它可以通过人称和句式替换来重复用户的句子，例如用户说“我感到孤独和难过”，ELIZA 会说“为什么你感到孤独和难过”。

这一系列的 AI 技术，让许多人对未来充满了憧憬。

当时有很多学者认为：“二十年内，机器将能完成人能做到的一切。”

没想到，事与愿违，AI 的发展很快陷入了停滞。

1.2.2　符号主义的低谷（1965—1970 年）

在达特茅斯会议上，纽厄尔和司马贺的成果激励了同行们，也激励了他们自己。

1957 年，纽厄尔和司马贺又弄了个“通用问题解决器”（General Problem Solver，GPS），听名字就知道他们期望用这个程序去解决任何已形式化的、具备完全信息的问题，包括逻辑推理、定理证明到人机游戏对弈等多个领域。这个项目持续了十多年，但并没有哪个领域有直接使用这个程序完成证明的大案例，可见并不成功。

在这个阶段的探索验证了一个观点，无论是逻辑理论家、通用问题求解器，还是几何定理证明机，这些发明都只适用于简单问题的求解，一旦涉及更难的问题选择时，结果证明都非常失败。

以机器翻译为例，准确的翻译需要背景知识来消除歧义，并建立句子内容。人类之所以了解背景知识是在建立在持续学习的基础上，但当时 AI 并不具备主动自我学习能力。到 1966 年，机器翻译被最终定性为“尚不存在通用科学文本的机器翻译，近期也不会有”。

为什么会这样？

因为符号主义的逻辑推理方法有极为明显的优缺点。优点是清晰简明，易于理解和实现，缺点是难以应对大规模、复杂的问题，并且难以处理不确定性和模糊性。事实证明，当需要证明包含超过数十条事实的推理时，原先的程序原则就失效了。

于是 AI 开始遭到批评，有人说所有的 AI 程序都只是“玩具”，仅仅能够解决诸如迷宫、积木世界这些“玩具问题”，投入产出比实在不划算。因此到 20 世纪 60 年代末，各类研究资助大幅缩减，只保留了一些基础研究。

其实，对解决实用的 AI 问题来说，这个阶段的 AI 理论、数据和算力都不够。

数据、算力不够没办法，只能等，但在理论方面，AI 开始形成了理论体系。

1969 年，明斯基被授予图灵奖，是历史上第一位获此殊荣的 AI 学者，因为他在 AI 的多个

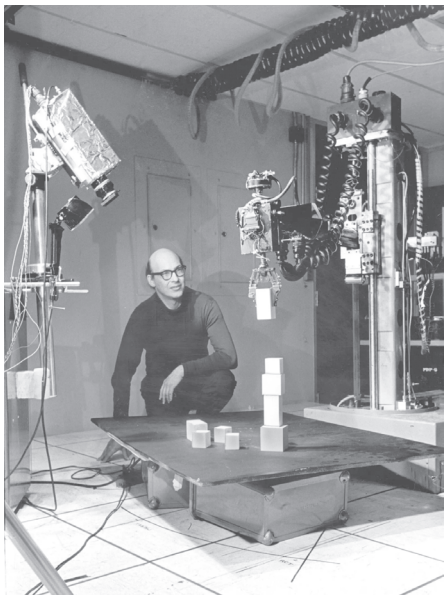


图 1-10　1968 年，明斯基在 MIT 的实验室（来源：nytimes.com）

领域如机器视觉、自然语言理解、知识表示和机器学习等方面都有杰出贡献。他还设计并制造了带有扫描仪和触觉传感器的机械手，可以像人手一样灵活地搭积木（如图 1-10 所示）。

明斯基认为，AI 的核心问题是如何使用符号来表示和操作知识，并通过逻辑推理和规则处理来模拟人类的思维过程。

1975 年，明斯基在 AI 领域提出了一种叫作“框架理论”的理论和方法。框架就像是一个表示知识的“盒子”，里面包含了固定的概念、对象或事件。盒子有多层，每层有很多小“格子”可以填入不同的信息，比如某个对象的特征、属性、取值范围等，这些“格子”也可以互相关联，组成一个完整的知识系统。

框架理论可以用于解决自然语言处理、机器学习和计算机视觉等领域中的问题。在自然语言处理中，框架理论可以用来构建语义解析器，帮助计算机理解自然语言中的语义信息；在计算机视觉中，框架理论可以用来建立视觉场景的模型，帮助计算机理解视觉场景中的物体、人物等概念。

除此之外，框架理论还为 AI 领域中的其他技术和理论提供了启示和指导，比如基于知识的推理、本体论和元认知等。

1.2.3　符号主义的“知识期”（1970—1980 年）

经过总结经验教训，研究者认识到知识在 AI 中的重要性，开始研究如何将知识融入 AI 系统。于是 AI 的发展阶段进入了知识时代，知识库系统和知识工程成为这个时期 AI 研究的主要方向。

到了 20 世纪 70 年代，研究者意识到智能的体现并不能仅靠推理来解决，知识是更重要的因素，研究重点就转变为如何获取知识、表示知识和利用知识，这个时期称为符号主义的“知识期”。在这一时期，出现了各种各样的专家系统，并在特定的专业领域取得了很多成果，让 AI 直接创造了商业价值，又一次推动了 AI 的发展。

知识工程最早由斯坦福大学的爱德华·费根鲍姆（Edward Albert Feigenbaum，如图 1-11 所示）提出。他与遗传学教授、诺贝尔奖获得者莱德伯格（Laideboge Joshua Lederberg）、布鲁斯·布坎南（Bruce G.Buchanan）等合作开发了世界上第一个专家系统 DENDRAL，该系统能够自动做决策，帮助化学家判断某待定物质的分子结构，目的是研究假说信息，构建科学经验归纳模型。

图 1-11　爱德华・费根鲍姆（来源：alchetron.com）

因在专家系统、知识工程等方面的贡献，费根鲍姆于 1994 年获得图灵奖。

什么是专家系统？

简单来说，专家系统可以根据某个领域已有的知识和经验进行推理和判断，最终做出模拟人类专家的决策，从而解决需要人工判断的问题。相比搜索，专家系统的出现让 AI 开始具备决策能力。

1976 年，一个名叫“MYCIN”的医疗专家系统问世，它可以帮助医生对患有血液感染的患者进行诊断和治疗。这个系统是依据专家们的想法建立的，用符号来表达治病症状和治疗方法等，还可以通过问题与回答的方式来推断病人是否患有某种疾病。MYCIN 内部有很多规则，只要按照程序的要求回答，系统就能判断病人感染的病菌种类，然后推荐给医生给病人吃的药。MYCIN 的处方准确率是 69%，虽然比不上专家医生的准确率 80%，但已经比其他不是医生的人做得更好了。

专家系统可以简单理解为“知识库＋推理机”，是一类具有专门知识和经验的计算机智能程序系统。因为专家系统一般采用知识表示和知识推理等技术来完成通常相关领域专家才能解决的复杂问题，所以专家系统也被称为基于知识的系统。

1980 年，卡内基梅隆大学为 DEC（Digital Equipment Corporation，数字设备公司）设计出了一套专家系统——XCON，它运用计算机系统配置的知识，依据用户的订货，选出最合适的系统部件，如中央处理器的型号、操作系统的种类及与系统相应的型号、存储器和外部设备以及电缆的型号，并指出哪些部件是用户没涉及而必须加进的，以构成一个完整的系统。它给出一个系统配置的清单，并给出一个这些部件装配关系的图，以便技术人员装配 DEC VAX 计算机。XCON 获得了巨大的成功，该系统在 DEC 公司内使用时，系统的规则从原来的 750 条发展到 3000 多条，最高时每年为公司省下 4000 万美元，这是第一个实用商用并带来经济效益的专家系统。

看到这样的前景，全世界的很多公司开始研发和应用专家系统。美军在伊拉克战争中也使

用了专家系统为后勤保障做规划，据称在 20 世纪 80 年代已有约 2/3 的美国 1000 强企业在日常业务中使用了专家系统。

但是，随着专家系统的应用领域越来越广，问题也逐渐暴露了出来。比如，XCON 专家系统的维护费用居高不下，难以升级，难以使用，当输入异常时会出现莫名其妙的错误。

而且，建立专家系统时也有很多问题。

首先，知识库的建立需要大量时间、人力和物力。一个专家系统能否成功，很大程度上取决于是否足够有效地整理了专家知识。困难在于，领域专家一般不懂 AI，专家系统的构建者也不懂领域知识，双方沟通起来非常困难。

其次，有些模糊无法定量的问题难以用文字衡量。专家可以解决某个问题，但很多情况下，专家又难以说清楚在具体解决这个问题的过程中，运用了哪些知识，因此知识获取成为建造专家系统的瓶颈问题。如果不能有效地获取专家知识，那么建造的专家系统就没有任何意义。

再次，知识 / 信息是无限的。如果知识仅依靠人类专家总结、提炼然后输入计算机的方式，则无法应对世界上几乎无穷无尽的知识。

最后，机器翻译始终无法取得预期效果，也影响了专家系统的跨语言使用。

这些问题在当时基本无解，导致专家系统的应用范围很有限。到了 20 世纪 80 年代后期，许多上马专家系统的公司发现掉进了大坑，系统难以升级，最终沦为只能解决某些特殊情景问题的处理机，AI 进入了第二个寒冬。

1.2.4 符号主义的“学习期”（1980—1990 年）

尽管专家系统为 AI 的发展定下了通过知识进行决策的基调，但这一路径的最大障碍是——知识从哪里获取？

于是，研究者们的重点又转为如何让机器自己学习知识、发现知识这个方向，这个时期被称为符号主义的“学习期”。

符号主义流派在“学习期”研究了更加灵活的符号主义机器学习算法。

这个时期，科学家们开始研究如何用计算机来学习各个领域的知识。他们希望机器可以像人一样学习，所以他们使用各种符号代表机器语言，来描述和组织这些知识，同时运用图表和逻辑结构方面的知识来设计机器。与此同时，科学家们探索了不同的学习策略和方法，并使用它们从数据中提取知识，比如，机器学习的图形表示，将低级别特征转化成更高级别的特征，为符号处理提供了更加精确的基础。

20 世纪 80 年代以来，被研究最多、应用最广的是“从样例中学习”，即从训练样例中归纳出学习结果，也就是广义的归纳学习。一大主流是符号主义学习，其代表包括决策树和基于逻辑的学习。

在将学习系统结合到各种应用中后，取得了很大的成功，同时专家系统在知识获取方面的需求也促进了机器学习的研究和发展。

20 世纪 80 年代是机器学习成为一个独立的学科领域，各种机器学习技术百花初绽的时期。

但在机器学习的研究过程中，人们发现符号主义存在一些无法解决的问题，包括：

（1）不能很好地处理不确定性和模糊性方面的情况。在很多场景下，往往是相关性规则，没有明确的因果性，而且变量太多了以后根本处理不过来。

（2）缺乏实际应用。因为需要大量的手工编码，难以适应新的情景和问题，比如图像和声音这种多媒体的内容，根本无法规则化。

（3）缺少灵活性。符号主义方法一旦确定了一个体系结构和规则，就难以适应新的情况和需求。

这些问题几乎都是无解的。

就这样，图灵倡导的符号主义理论体系，由四位获得了图灵奖的 AI 之父们带路，最终居然走进了死胡同。

这可怎么办？好在，人类在 AI 方面的探索并不止一条路，有其他科学家找到了新路。

1.3　连接主义流派的探索

连接主义（Connectionism），也叫联结主义，又称“仿生流派”或“生理流派”。连接主义认为，智能不是由一组规则或知识表示所构成的，而是由大量简单元素互相作用而形成的。它提出了“连接权重学习”的概念，即通过权重的自适应学习来更好地适应和解决问题。与符号主义不同，连接主义更适合处理不确定性和模糊性等复杂情况，因为它具有适应性学习模式，能够不断地从反馈信息中学习、调整自身的行为或表现，以适应不同的情境和需求，所以被广泛应用于语音识别、图像识别、自然语言处理等领域。

这个流派的思想其实很简单，既然人类的智能是大脑的活动所产生的结果，机器智能也像大脑一样，用大量简单的单元通过复杂的相互联结后并行运行产生智能。这符合第一性原理。

人脑中的神经元是一种特殊的细胞，有着很多长长的、细小的突触，通过这些突触与其他神经元进行联系和沟通。

当受到刺激时，神经元内部的电荷会发生改变，形成一个“动作电位”。这个动作电位会被传递到神经元的轴突末端，然后释放出一些化学物质，称为神经递质，来刺激与它相连的其他神经元。

如果这个被刺激的神经元与足够多的其他神经元联系起来，它们就可以形成一个神经网络，这些复合的信号就是我们思考、感觉和行动的基础。

人工神经网络借鉴了人脑的神经网络，其中的每个神经元都与其他神经元相连，这些连接具有不同的权重，可以修改以适应不同的问题。神经网络使用并行分布式处理的方式，可以实现模式识别、分类、预测等多种模式匹配任务。

一般人脑有 120 亿～140 亿个神经元，科学家们发现至少需要一个 5 层神经网络才能够模

拟单个生物神经元的行为，也就是说需要大约 1000 个人工神经元才能够模拟一个生物神经元。GPT-3 有 1370 亿个参数，也就相当于模拟了人类的 1.37 亿个神经元，相当于人脑的 1%。有人预测，如人工神经网络的模拟成果达到了人脑的神经元数量，就能通过图灵测试。

如今 AI 的主要流派是连接主义，使用神经网络来模拟人类大脑。神经网络学习方法的基本原理就是从用于训练的数据集中提取出分类特征，这些特征应能同样适应独立同分布（Independent Identically Distribution）的其他未知数据，所以经已知数据学习训练后的神经网络可以对同类的未知数据有效。

但这个流派的发展却一波三折，一开始就不被看好，红火了一阵后，又被明斯基的一本书搞得停滞了近 20 年，然后再次兴起，最终成为主流。

1.3.1 人工神经元的探索（1943—1967 年）

1943 年，数学家沃尔特·皮茨（Walter Pitts）和神经生理学家沃伦·麦卡洛克（Warren McCulloch）发布了一篇划时代的论文《神经活动中内在思想的逻辑演算》（A Logical Calculus of the Ideas Immanent in Nervous Activity），提出了生物神经元的计算模型“ M-P 模型”，并进行了神经元的数学描述和结构分析。后来，普林斯顿大学高级研究员的冯·诺伊曼从麦卡洛克 - 皮茨模型中获取了灵感，提出了“冯·诺伊曼结构”。1945 年 6 月，冯·诺伊曼在划时代的论文《EDVAC 报告书的第一份草案》（*First Draft of a Report on the EDVAC*）中唯一引用公开发表的文章，就是麦卡洛克和皮茨这篇论文。由于发明了第一台电子数字计算机 EDVAC，冯·诺伊曼被称为计算机之父。

麦克洛克认为，大脑的工作机理很可能是这样的一种机器，它用编码在神经网络里的逻辑来完成计算。如果神经元可用逻辑规则连接起来，那就构建了结构更为复杂但功能更加强大的思维链（chains of thoughts），这种方式与《数学原理》将简单命题链（chains of propositions）连接起来以建构更加复杂的数学定理是一致的。皮茨用数学证明了只要有足够的简单神经元，在这些神经元互相连接并同步运行的情况下，就可以模拟任何计算函数。

这篇论文将神经网络抽象成了数学模型，提出了阈值逻辑单元（threshold logic units，TLU）这个函数模型来描述神经元，并用环形的神经网络结构来描述大脑记忆的形成。两位作者认为，随着长时间对神经元阈值的调整，随机性会渐渐让位于有序性，而信息就涌现出来了。这个预测已经被 77 年后的 GPT-3 语言模型所证明，当模型参数达到千亿规模的时候，它展现出了令人惊叹的智能，出现了涌现效应。

不过在那个年代，完全没有条件去规划调整神经元阈值，也没什么合适的理论指导。在黑暗中摸索太难了，早期的研究没什么拿得出手的成果。

明斯基在读博士期间发明的“SNARC”机器，可算是世界上第一批基于神经网络的自学习机器的工程实践成果，不过这个成果在 1956 年的达特茅斯会议上没有得到大家的好评，当时的主要研究者都看好符号主义的发展方向。后来明斯基自己也“叛变”了，兴趣转移到其他发展方向上。

直到 1957 年，美国海军公布了一个实验，一下子让大众都对神经网络兴奋了起来。

这个实验看起来很简单，一台足有 5 吨重、面积大若一间屋子的 IBM 704 被“喂”进一系列打孔卡，经过 50 次实验，计算机自己学会了识别卡片上的标记是在左侧还是右侧。

就这么个无趣的实验，意义却很重大。它展示了一条全新的实现机器模拟智能的道路——不依靠人工编程，仅靠机器学习就能完成一部分机器视觉和模式识别方面的任务。

这个技术的创始人弗兰克·罗森布拉特（Frank Rosenblatt）说：“创造具有人类特质的机器，一直是科幻小说里一个令人着迷的领域。但我们即将在现实中见证这种机器的诞生，这种机器不依赖人类的训练和控制，就能感知、识别和辨认出周边环境。”

罗森布拉特的说法引起了媒体和新兴的 AI 学界的浓厚兴趣。《纽约时报》将《海军的新设备可以通过实践来学习：心理学家展示了一种在阅读中变聪明的计算机雏形》用作头条标题，而《纽约客》写道：“它确实是人类大脑的第一个正经的竞争对手。”

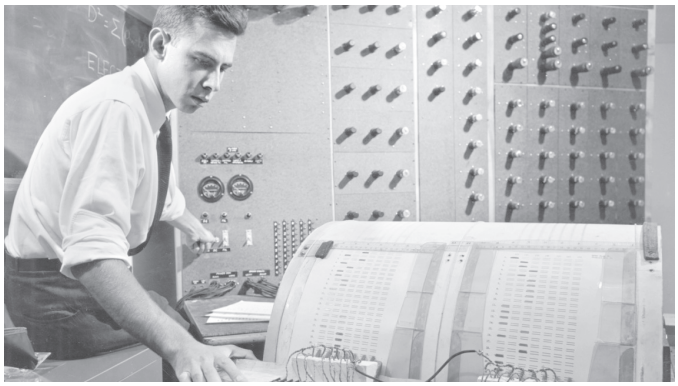


图 1-12　弗兰克·罗森布拉特在“感知机”上工作（来源：康奈尔大学官网）

罗森布拉特是康奈尔大学的实验心理学家，他模拟实现的这个装置是一个叫作“感知机”（Perceptron）的神经网络模型（如图 1-12 所示）。于是政府拨款几十万美元，大力支持罗森布拉特的神经网络研究，罗森布拉特又做了多个关于感知机学习能力的实验。

1962 年，罗森布拉特出了一本书，书名叫《神经动力学原理：感知机和大脑机制的理论》（*Principles of Neurodynamics：Perceptrons and the Theory of Brain Mechanisms*），此书总结了他对感知机和神经网络的主要研究成果，证明了单层神经网络在处理线性可分的模式识别问题时是可以收敛的，一时被连接主义流派奉为“圣经”。

看来，连接主义的发展开了一个好头。

没想到，明斯基却站出来，提出尖锐的反对意见。

他们俩是高中校友，隔一年毕业，难道是私人恩怨？

你想多了，纯粹是学术之争。

罗森布拉特认为，他可以用感知机技术让计算机阅读并理解语言，而明斯基说不可能，因为感知机的功能太简单了。

虽然明斯基在其后期工作中也关注了一些连接主义的元素，但他仍然认为符号主义是解决 AI 核心问题的最佳方法。

明斯基在一次会议上和罗森布拉特大吵了一架，随后，明斯基和 MIT 的另一位教授西摩尔·派普特（Seymour Papert）合作，企图从理论上证明他们的观点。

在明斯基看来，罗森布拉特的神经网络模型存在两个关键问题。

首先，单层神经网络只能处理线性的数据，无法处理“异或”电路，也就不能处理非线性的数据问题。举个例子，假设要训练一个神经网络来判断水果是不是橙色的，单层神经网络只能够处理线性数据，也就是只能判断水果是不是红色的或者黄色的，而不能够判断是不是橙色的。如果使用多层神经网络，就可以将红色的苹果、黄色的香蕉和橙色的橘子都正确分类，因为多层神经网络可以组合非线性函数，从而处理复杂的非线性问题。这就好比把红色和黄色的果汁混合在一起，就能够得到橙色的果汁一样。所以多层神经网络就像是一个果汁机，可以把不同颜色的果汁混在一起得到各种不同的口味。

其次，虽然多层神经网络可以通过多个层级的组合和非线性操作来表现各种数据特征，但当时计算机的算力不够，无法满足多层感知机构成的大型神经网络长时间运行的需求。

1969 年，明斯基和派普特（如图 1-13 所示）合著的《感知机：计算几何学导论》（*Perceptrons: An Introduction to Computational Geometry*）一书出版，给了罗森布拉特致命一击。

图 1-13　马文·明斯基（左）与西摩尔·派普特（右）（来源：MIT 官网）

由于明斯基在 AI 领域中的特殊地位，再加上他不久前刚获得图灵奖所带来的耀眼光环，这本书不仅对罗森布拉特本人，还对连接主义和神经网络的研究热情，甚至对整个 AI 学科都造成了非常沉重的打击。

明斯基还给出了他对多层感知机的评价和结论：“研究两层乃至更多层的感知机是没有价值的。”也就是说，他认为连接主义的神经网络这条路是绝路。神经网络和深度学习技术的研究迅速陷入了停滞，大部分人都退出了，用加州理工学院的集成电路大佬米德（Carver Mead）的话说是“20 年大饥荒”。

当时没有任何人能够预见到，神经网络和连接主义在 20 年后还会有机会逆袭，最终成为 AI 研究的最主流最热门的技术。

如果时光倒流，明斯克没有打击多层感知机的研究，会出现什么呢？

我们以识别猫来做一个思想实验，就像罗森布拉特识别卡片一样，某个专家利用神经网络

来训练电脑去识别猫。先让电脑学会观察照片，找出其中独特的特征，这些特征可能包括颜色、亮度、轮廓和纹理等。如果电脑只看过 5 张猫咪图片，那就只有 5 个拍摄视角、光线环境，肯定不够总结出足够的特征，但如果电脑看过 500 张猫咪图片，那就可以有更多的例子来得出它们之间的共性。然后发现还不够，最终需要收集几万张图片来进行训练迭代，让它逐渐提高识别准确度。每一次训练后，电脑就会将它学习到的特征存储在一个名为“特征”的列表中。最终它会通过分析新照片中的这些特征，来判定这张照片中是否有猫存在。这个多层感知机的研发路线是对的，但以当时的条件根本做不到。在 20 世纪 60 年代的计算条件下，算力根本做不到足够多的训练，也就不能从照片上识别猫。

人类是什么时候解决了 AI 识别猫的问题的呢？

直到 40 多年后，2012 年吴恩达参与领导的 Google Brain（谷歌大脑）运用 16000 个超级计算机训练深度神经网络，最终让计算机成功识别出猫，成为深度学习与产业结合的轰动性事件，而这个神经网络有数百万层。

所以，明斯基没说错，如果按罗森布拉特的感知机路线继续走下去，以当时的条件，只是浪费资金，不可能出成果。

但罗森布拉特的构想确实开启了新的路线，这对神经网络的后续发展非常重要。2004 年，美国电气电子工程师学会（Institute of Electrical and Electronics Engineers，IEEE）设立了 IEEE 弗兰克 • 罗森布拉特奖（Frank Rosenblatt Award）。

遗憾的是，罗森布拉特没有等到这一天。

1971 年 7 月 11 日，也就是罗森布拉特 43 岁生日那天，他在切萨皮克湾划船时，离奇地溺水身亡了。这实在是太可惜了。

1988 年，明斯基对《感知机：计算几何学导论》一书进行改版修订时，删除了攻击罗森布拉特个人的句子，并在第一页手写了“纪念弗兰克 • 罗森布拉特”（In memory of Frank Rosenblatt，如图 1-14 所示）。

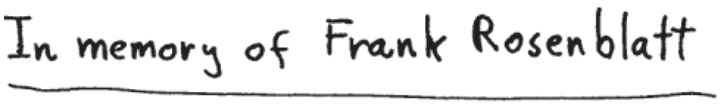


图 1-14　明斯基手写的话语（来源:《感知机：计算几何学导论》电子书）

1.3.2　第二代神经网络（1982—1995 年）

山重水复疑无路，柳暗花明又一村。20 世纪 80 年代末，神经网络的研究迎来了第二次兴起，这源于分布式表达与反向传播算法的提出，减少了对算力的要求；同时，计算机技术的飞速发展也使计算机有了更强的计算能力，此时的计算成本几乎仅为 1970 年的千分之一，这些因素使神经网络解决了明斯基提出的尖锐问题。于是，被明斯基说成绝路的连接主义路线，又打开了新迷宫的关卡。

打开这个新关卡，最需要感谢的有五个人：约翰·霍普菲尔德、杰弗里·辛顿、杨立昆、约书亚·本吉奥和于尔根·施密布尔，他们提供了必要的核心组件。其中，辛顿被誉为“深度学习教父”，正因为以他为首的一批坚定信仰神经网络的学者们的共同努力，最终使得神经网络能够以崭新的面貌登上人工智能舞台，并获得主角地位。下面介绍一下他们的贡献。

图 1-15　约翰·霍普菲尔德
（来源：alchetron.com）

1. Hopfield **模型**

1982 年，生物物理学家约翰·霍普菲尔德（John Hopfield，如图 1-15 所示）发布了 Hopfield 模型。Hopfield 模型被广泛运用于神经学领域，如研究脑细胞如何进行信息处理和存储，也为人工神经网络提供了基础理论的支持。

Hopfield 模型可以对存储的信息进行联想记忆，它可以根据输入的信息自动产生关联并输出结果，因此也可以用于模式识别和任务分类。例如，用 Hopfield 模型可以存储和检索人脸图片，也可以通过训练来识别不同的人脸。

在 Hopfield 模型中，神经元通过相互作用来达到一种稳定的状态，以表示特定的信息，还可以在有噪音和不完整信息的情况下存储和检索信息，这种稳定状态的特性使 Hopfield 模型非常适合用于分类、压缩和优化等领域。

2001 年，约翰·霍普菲尔德因这个贡献获得了理论和数学物理领域的最高荣誉——ICTP 狄拉克奖，Hopfield 模型也被视为人工神经网络重启的里程碑。

2. **反向传播算法**

1986 年 7 月，杰弗里·辛顿（Geoffrey Hinton）和他的学生大卫·鲁姆哈特（David Rumelhart）共同发表论文，提出反向传播算法，这个算法的伟大之处在于大大减少了运算量。传统的感知器用所谓“梯度下降”的算法纠错时，其运算量和神经元数目的平方成正比，而这个算法把纠错的运算量下降到只和神经元数目成正比。这就相当于，当有 10 万个神经元的时候，传统算法需要 100 亿次计算，而新算法只需要 10 万次计算。

增加隐藏层（hidden layer）是该算法的特点，它引入了非线性结构，解决了第一代神经网络的线性不可分问题。因此，该算法能够处理非线性问题，包括明斯基提出的感知器无法解决的异或门难题。

这篇论文奠定了人工智能领域神经网络的基础，科学家终于可以组建更庞大的人工神经网络来处理更多更复杂的问题了。

为了取得这个成就，辛顿付出的艰辛鲜为人知。

辛顿第一次听说神经网络是在 1972 年，当时他在爱丁堡大学攻读人工智能专业硕士学位，因为他在剑桥大学读本科时研究的是实验心理学，所以对于神经网络很有热情。

辛顿认为，神经网络是一种比逻辑思维更为优秀的智能运作模式。这是因为神经网络能够

让计算机像人一样学习。

那时，神经网络已经进入了寒冬期，而杰弗里·辛顿却在爱丁堡大学攻读神经网络博士学位。

当时他在论文中只要提到“神经网络”，论文就无法通过同行评审。

有人对辛顿说：“这太疯狂了。你为什么要在这些东西上浪费时间？事实已经证明这是无稽之谈。”

他回答说，即使他没有获得博士学位也没关系。辛顿出生在英国一个有着深厚学术内涵的家庭，除了科学家，他从未考虑过其他任何事情。辛顿知道，“一些流行的智慧是无可救药的错误——但不一定是哪一点”。

辛顿 1978 年完成博士学位后，四处走动，寻找研究天堂。他在加州大学圣地亚哥分校做博士后，度过了一段至关重要的时光。他说，那里的学术氛围让人更容易接受，并且他在那里与认知神经科学先驱大卫·鲁姆哈特进行了合作。

回到英国后，辛顿做着一份无聊的工作。20 世纪 80 年代初，辛顿和同事开始尝试让电脑模仿大脑运作，但当时电脑性能还远远不能处理神经网络需要的巨大数据集，获得的成果很少。AI 社区的支持者们也放弃了，转而去寻找类人脑的捷径。

一天半夜，辛顿被一个美国来的电话惊醒，对方表示，愿意资助他 35 万美元继续他的研究。

辛顿后来才知道这笔资助的来源：兰德公司的一个非营利子公司通过开发核导弹攻击软件获得了数百万美元。因为是非营利组织，政府要求他们，要么把这笔钱用来支付薪水，要么尽快赠予研究机构，他们选择把这笔钱送给了辛顿。

因为这笔钱，辛顿等人的研究又复活了。

1986 年，辛顿迁居加拿大，在多伦多大学任教，并加入了加拿大高等研究院（CIFAR），获得了一个能长时间从事基础研究的职位，参与了首个名为“AI、机器人与社会研究”的项目（如图 1-16 所示）。

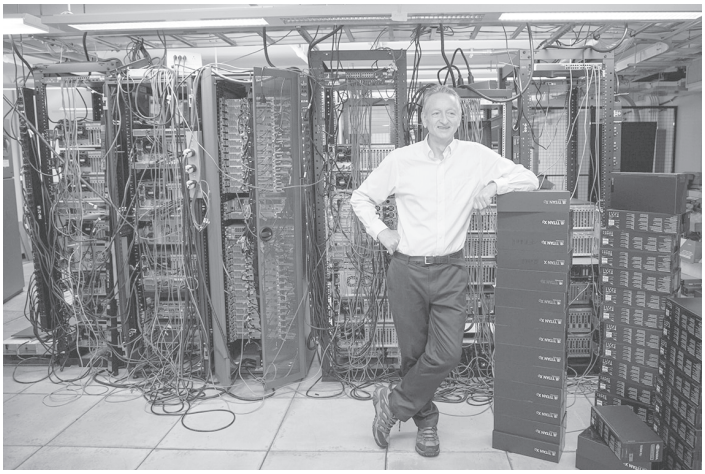


图 1-16　辛顿在多伦多大学的机房（来源：多伦多大学官网）

辛顿终于安顿了下来，在加拿大逐渐有了核心团队，最终收获了梦寐以求的成果。

辛顿有一个习惯，喜欢突然大喊："我现在理解大脑是如何工作的了！"这很有感染力，他常常这样做。

在漫长的神经网络寒冬期，怕是只有这样的自我激励才能坚持下去吧！

3. 卷积神经网络

图 1-17　杨立昆（来源：UCLA 官网）

1989 年，法裔美国计算机学家杨立昆（Yann LeCun，如图 1-17 所示）提出了卷积神经网络，这项研究成果是 AI 深度学习历史上划时代的里程碑。卷积神经网络是一种深度学习模型，它的特点是可以自动从数据中学习特征，并且在处理图像、语音、自然语言等任务时表现出色，已经成为深度学习中几乎不可或缺的组成部分。

杨立昆曾在加拿大多伦多大学跟随辛顿教授做博士后研究，所以算辛顿的半个学生。1988 年，杨立昆进入贝尔实验室，在那里他和一位名叫约书亚·本吉奥（Yoshua Bengio）的博士后使用神经网络进行数字识别。

做这个项目的时候，他们发现了当时神经网络的一个大问题：由于图像中包含的像素数量非常庞大，相应的神经网络的参数量也很大，这在很大程度上影响了手绘图片的识别准确率，为解决此问题，他们想到一个巧妙的算法。

我们用侦探来类比这个算法，神经网络需要根据线索（即输入数据）来推断出真相（即输出结果）。侦探只关注案件中的某些细节，而不是整个案件。卷积神经网络只对输入数据的一部分进行处理，而不是对整个数据进行处理。类似侦探可以使用同样的技能在不同的案件中寻找线索，卷积神经网络的参数共享特征让每个卷积核都可以在输入数据的不同位置上使用。侦探可以从大量的线索中筛选出重要线索，卷积神经网络有池化层，它可以减小输入数据的数量，同时保留重要的特征。

他们用美国邮政系统提供的近万个手写数字的样本来训练神经网络系统，在独立测试样本中错误率低至 5%，达到实用水准。

之后，杨立昆进一步运用卷积神经网络技术开发出支票识别的商业软件，用于读取银行支票上的手写数字，这个系统在 20 世纪 90 年代末占据了美国接近 20% 的市场。

卷积神经网络用途非常广泛，在人脸识别、语音识别和图像分类等方面都能被用到。

2013 年，杨立昆以纽约大学教授的身份兼职加入脸书，随后便着手组建了脸书的 AI 实验室。杨立昆领导的脸书 AI 研究（Facebook AI Research，FAIR），主攻方向是自然语言处理、机器视觉和模式识别等。脸书还组建了"应用机器学习"（Applied Machine Learning, AML）部门，由华金·坎德拉（Joaquin Candela）带领，主要负责将机器学习的研究成果落地到产品上，与 FAIR 的资源分配比例正好互补。

得益于 FAIR 和 AML 的工作成果，脸书在自然语言处理和人脸识别方面确实处于业界领先地位，例如他们的人脸识别工具 DeepFace 已经能做到比人类更准确地识别出两个不同图像上的人是否是相同的。

4. 序列概率模型

在 1990 年，约书亚 • 本吉奥（Yoshua Bengio，如图 1-18 所示）在贝尔实验室做博士后的时候，跟杨立昆一起做数字识别的项目。他提出将神经网络与序列概率模型（例如隐马尔可夫模型）相结合，这个策略后来被纳入 AT&T/NCR 用于读取手写支票的系统中，并被认为是 20 世纪 90 年代神经网络研究的巅峰之作。本吉奥和杨立昆可以说是天作之合，在项目合作中各自获得了自己的学术成就，并培养出深厚的友谊。

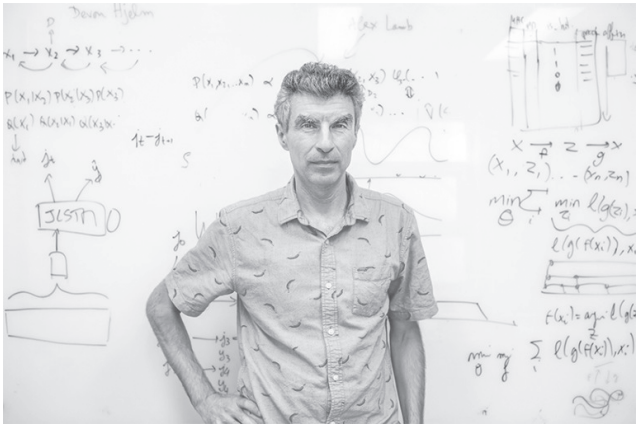


图 1-18　约书亚 • 本吉奥（来源：Nature.com）

什么是序列概率模型呢？

简单来说，序列概率模型是一种可以预测序列中下一个元素的模型。我们可以把序列概率模型比作一个预言家，它可以根据历史数据（即序列数据）来预测未来的趋势。

ChatGPT 就用到了序列概率模型。在生成文本时，ChatGPT 首先输入一个起始文本序列，然后使用序列概率模型分析下一个单词的概率分布。根据概率分布，ChatGPT 随机选择一个单词作为下一个单词，并将其添加到当前序列中。然后，ChatGPT 使用更新后的序列再次预测下一个单词，并重复这个过程，直到生成所需长度的文本。

比如对一个句子：not all heroes wear ___，ChatGPT 将会分析下一个单词出现的概率，并选择较大概率的单词。猜得越准，就越显得像人类。

2000 年，本吉奥发表了一篇具有里程碑意义的论文《神经概率语言模型》（*A Neural Probabilistic Language Model*），通过引入高维词嵌入技术实现了词义的向量表示，将一个单词表达为一个向量，通过词向量可以计算词的语义之间的相似性。该方法对包括机器翻译、知识问答和语言理解等在内的自然语言处理任务产生了巨大的影响，使应用深度学习方法处理自然语言问题成为可能，并使相关任务的性能得到大幅度提升。

2007 年，本吉奥负责蒙特利尔大学机器学习算法实验室时，他和伊恩 • 古德费洛（Ian Goodfellow）一起开发出了第一个开源的深度学习框架 Theano，这个框架启发了谷歌 TensorFlow 深度学习框架的开发。

2014 年，本吉奥与古德费洛提出的生成对抗网络（Generative Adversarial Network，GAN），引发了一场计算机视觉和图形学的技术革命，使计算机生成与原始图像相媲美的图像成为可能。它被誉为近年来最酷炫的神经网络。

本吉奥的团队还提出了注意力机制，直接导致机器翻译取得了突破性进展，并构成了深度学习序列建模的关键组成部分。ChatGPT 是基于谷歌 Transformer 模型的，而 Transformer 就是基于注意力机制推出的。

什么是注意力机制？注意力机制来源于人类的视觉注意力，即人类在进化过程中形成的一种处理视觉信息的机制。人类视觉系统以大约每秒 8.96 兆比特的速度接收外部视觉信息。虽然人脑的计算能力和存储能力都非常有限，却能有效地从纷繁芜杂的外部世界中有选择地处理重要的内容，在这个过程中选择性视觉注意力发挥了重要的作用，如我们在看一个画面时，会有一处特别显眼的场景率先吸引我们的注意力，这是因为大脑对这类东西很敏感。

注意力机制使得模型在翻译的过程当中并不以同等地位看待所有的词，而是根据当前翻译的需要着重关注少数几个词的信息。这样既能尽量保留一段文字中的各种细节，又能得到较为准确合理的翻译。

2017 年 1 月，本吉奥成为微软公司的战略顾问。他认为微软这个曾经的“Windows 帝国”可以将自己打造成 AI 第三巨头，微软有资源、有数据、有人才，还有愿景与文化（这是最重要的），它不但知道科学的重点在哪里，还推动着技术向前发展。

鉴于辛顿教授、杨立昆教授和本吉奥教授（如图 1-19 所示）三人对深度学习的贡献，2018 年三人同时获得图灵奖。

图 1-19　辛顿、杨立昆和本吉奥（从左至右）（来源：nytimes.com）

5. 长短时记忆循环神经网络

1997 年，瑞士 AI 实验室（IDSIA）的于尔根 • 施密布尔（Jürgen Schmidhuber，如图 1-20 所示）和塞普 • 霍克利特（Sepp Hochreiter）共同发表论文，提出了长短时记忆循环神经网络（Long Short-Term Memory，LSTM），为神经网络提供了一种记忆机制，可以有效解决长序列训练过程中梯度消失的问题。

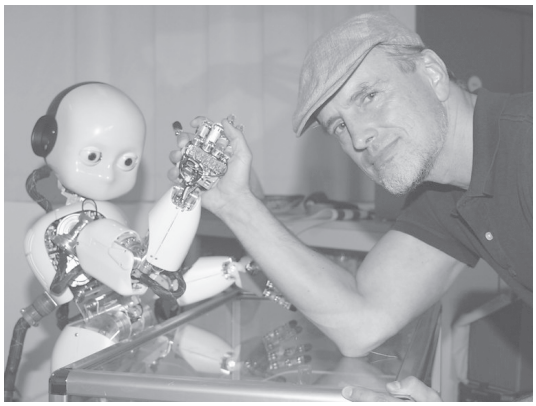


图 1-20　于尔根 • 施密布尔（来源：telekom.com）

简单理解 LSTM，可以将 LSTM 看成一位记忆超强、反应迅速的秘书，她可以处理不同类型的请求，并能够长期记住关键信息。当秘书接到一条请求时，她会仔细阅读，首先判断是否是关键请求，如果不是，她会将其处理后放入备忘录中；如果是关键请求，她会尝试记住关键内容，并暂时放在头脑中，等待未来可能的需要。当类似的请求再次出现时，秘书会立即回忆起之前的关键信息，并根据之前的经验进行相应的处理。

实践证明，这一技术在序列问题的处理，如自然语言理解和视觉处理中，发挥着至关重要的作用，并广泛应用于机器翻译、自然语言处理、语音识别和对话机器人等任务。以语音识别为例，每个音节作为输入数据，LSTM 对其进行处理和采样，生成音频信号的表示，然后传递给下一个神经元进行处理。神经元之间的连接和权重会不断调整，逐步学习和记忆之前的上下文信息，以此识别整个语音信号。

在 2016 年和 2021 年，施密布尔和霍克利特分别被授予了 IEEE 神经网络先驱奖。

因为上述专家们的贡献，AI 的第二波浪潮开始到来。

一时间，人工神经网络成为大家热议的名词，在 1991 年的电影《终结者 2》中，施瓦辛格扮演的机器人有一句台词：“我的 CPU 是一个神经网络处理器，一台会学习的计算机。（My CPU is a neural-net processor，a learning computer）”。

尽管神经网络的发展道路已经打通，也取得了不少进展，可这股浪潮仅仅持续了不到十年。到 1995 年前后，大家又开始对 AI 失去信心，因为训练人工神经网络太慢了。一位国内的 AI 研究者回忆，一直到 1998 年，他做研究生的时候，在当时的电脑上运行 AI 程序，单元不敢超过 20 个，而人类大脑的神经元数量是 120 亿个以上，可想而知当时的神经网络是多么弱小。

在神经网络发展受阻的同时，传统的机器学习算法取得了突破性的进展，机器学习社区转向了能更快、更好提供结果的基于规则的系统，而兴起了没多久的神经网络逐步被取代。

据统计，到 21 世纪初，全世界专门研究神经网络的研究人员数量不到六人。

神经网络的研究者除了继续优化算法，更多的时间只能等待。

等待什么？等算力发展和数据积累。

1.3.3 统计机器学习（1995—2012 年）

1. 数据和算力的增长，推动了统计机器学习

自 1995 年互联网开始发展以来，海量数据的涌现为神经网络提供了更大的发展机遇。如果没有互联网，大公司使用的图像数据集、视频数据集和自然语言数据集根本无法收集。例如，Flickr 网站上用户生成的图像标签一直是计算机视觉的数据宝库，YouTube 视频也是一座宝库，维基百科则是自然语言处理的关键数据集。

算力的增长，主要靠 CPU 和 GPU。

根据“摩尔定律”，每 18 个月芯片内晶体管的数量增加一倍，简单来说，就是芯片的运算能力增加一倍。从 20 世纪 70 年代一直到现在，“摩尔定律”持续了将近 50 年的时间。

今天同样大小的 CPU 芯片，比当年的运算能力至少高出 100 万倍。如英特尔的 Core i9-11900K，其时钟频率可以高达 5.3 GHz，配备了多个核心和超线程技术，可以同时执行数百万个指令，其性能是 Intel 4004 的数百万倍。

而 GPU 登上舞台，又把算力提升了一个数量级，因为 GPU 的并行处理能力比 CPU 快得多。一开始，GPU 用于显卡，能提供更好的游戏体验。2007 年，英伟达发布了面向软件程序员的框架 CUDA（计算机统一设备架构），任何人都可以运用 CUDA API 在 NVIDIA GPU 上进行通用计算（GPGPU），于是深度学习模型纷纷采用 GPU 来计算。

1997 年 5 月 11 日，IBM 的 AI“深蓝”战胜了国际象棋棋王卡斯帕罗夫（如图 1-21 所示），他曾雄踞世界棋王宝座 12 年之久。

图 1-21 卡斯帕罗夫败给 IBM 深蓝（来源：wired.com）

IBM 深蓝的胜利，可以说是由算力发展带来的胜利。通俗点说，就是深蓝“背”的棋谱比人类多，脑子比人类转得快，于是人类就输了。

深蓝是一台重达 1.4 吨的超级计算机，因为有 32 个节点可以并行计算，每秒可以算出 2 亿步，完全可以按规定的时间每 3 分钟内从储存的棋谱中寻找出下一步该走的棋。

研究人员把将近 100 年来 60 万盘高手的棋谱都储存在“深蓝”的“外脑”——大型快速阵列硬盘系统中。“深蓝”系统是由两个数据库组成的：一个是开局数据库，它最初几步棋的下法都是到大约 2 兆字节的开局数据库中寻找的；另一个是终局数据库，数据量达到了 5 千兆字节。

IBM 深蓝的成功，让公众一下子对 AI 有了巨大的热情。

在深蓝的开发过程中，IBM 团队使用了强化学习中的“策略网络”技术，来生成深蓝的下棋策略。策略网络是一个神经网络，它接受下棋历史数据作为输入，并输出下棋的最佳策略。深蓝通过不断与环境交互，通过奖励和惩罚来训练策略网络，使网络能够逐渐逼近最优解。

此外，深蓝还使用了强化学习中的智能体（agent）技术，来模拟国际象棋游戏中的棋子。智能体通过与环境交互，不断学习和优化自己的策略，以提高下棋水平。

什么是强化学习呢？

简单来说，强化学习就是让 AI 每一步的学习都要获得反馈。比如 AlphaZero 下棋，它每走一步棋都要评估这步棋是提高了比赛的胜率，还是降低了胜率，以获得一个即时的奖励或惩罚，从而不断调整自己。

机器怎么知道如何评估对和错呢？

为了让机器学习的方向符合人类的期望，人类会对学习材料做一些标记，再“喂”给 AI 进行训练，让神经网络在训练过程中有的放矢，这就是监督学习（supervised learning）。监督学习的 AI 就好像学校里老师对学生的教学，对错分明有标准答案，但是可以不讲是什么原理。举个例子，就好像人类设计了一个规则，教 AI 玩一个迷宫游戏，以找到到达迷宫终点的最短路径为胜利。在这个游戏中，AI 每次走一步，就会得到或失去一些分数。如果 AI 走的是正确的路，就会获得更多的分数，但如果 AI 走的是错误的路，就会失去一些分数。通过不断地试错和调整，AI 最终可以找到一条通往终点的最短路径。

强化学习属于行为主义流派，仿照了人类从周围环境中不断获得反馈信息而学习的方式。

例如，当学生解决一个问题或完成一项任务后，会有老师给予反馈（比如做对或做错、获得奖励或惩罚等），学生会根据这些反馈信息来改进自己的学习行为。再比如，父母教孩子骑自行车，一开始，孩子可能会不断尝试调整身体的平衡，然后学习如何踩踏板推动车子前进，如果孩子成功骑到了几米远，父母通常会赞扬和鼓励他们，这种学习叫作强化学习，赞扬就是学习中的“奖励机制”。经过反复练习，并在大脑中建立新的神经元连接，最终，孩子形成了骑车技能，以后骑车就不用费心思把握平衡了。

行为主义的起源可以追溯到控制论，但直到 20 世纪末才被正式提出，其核心理念是智能依赖于感知、行为和对外部环境的适应能力。这种方法的优点是可以解决在环境和任务变化时的自适应性问题，广泛地应用于游戏、自动驾驶、机器人和智能机械等领域，比如波士顿动力机器人就是用到了大量行为主义的智能感知技术和训练方式。行为主义的强化学习也有缺点，需要大量的训练数据和较长的学习时间，这也是人形机器人发展速度不够快的一个原因。

因为行为主义出现的时间较短，就不展开讲行为主义流派的探索过程了。

总之，由于计算机性能的提高和大数据的出现，统计机器学习成了主流的 AI 学习方式，比如亚马逊、谷歌、奈飞、字节跳动等许多互联网公司用到的商品推荐系统、影视剧和内容推荐系统等，都属于统计机器学习范畴。

而辛顿倡导的神经网络在 AI 的研发领域却依然处于边缘化的地位，因为这个技术还没真正解决好学习的效率问题，实验室里的训练结果离达到商业化成果还太远。

全世界只有几个人还在坚持研究神经网络，但他们既缺少数据，也缺少算力，又能做出什么成果呢？

也许，他们在等待一个机会吧。

2. 深度学习诞生

由于找不到合适的经费来源，辛顿辗转于瑟赛克斯大学、加利福尼亚大学圣地亚哥分校、卡内基梅隆大学和英国伦敦大学等多所大学工作，2004 年他终于从加拿大高等研究院（Canadian Institute For Advanced Research, CIFAR）申请到了每年 50 万美元的经费支持。加拿大高等研究院可能是那个时候唯一还在支持神经网络研究的机构，现在看来这是一笔投入收益比惊人的投资。辛顿当时申请了研究课题为“神经计算和适应感知”的项目，虽然每年 50 万美元是一笔微薄的经费，但还是让辛顿在加拿大多伦多大学安顿下来，结束了飘摇不定的访问学者生涯。这样少得可怜的经费，比起其他知名的 AI 项目的巨额投资来说，简直就是讽刺。但辛顿决定要创建一个世界级的团队，致力于开发模拟生物智能的模拟程序——模拟人类大脑如何通过筛选大量的视觉、听觉和书面信息来理解和适应环境。

在杨立昆和本吉奥的支持下，辛顿建立了神经计算和自适应感知项目，这个项目还邀请了一些计算机科学家、生物学家、电气工程师、神经科学家、物理学家和心理学家。辛顿认为，建立这样一个组织会刺激 AI 领域的创新，甚至改变世界。在当时，他这样想是过于乐观了，但事实证明，他是对的。

两年后，辛顿的团队取得了突破，发表论文《深度信念网络的一种快速学习算法》（*A Fast Learning Algorithm for Deep Belief Net*），首次提出了贪婪逐层训练深度神经网络的方法，大大降低了深层神经网络的训练难度。这个深度信念网络（Deep Belief Network）被冠以新名称，即“深度学习”。

这个神经网络深度学习中的深度指的是神经网络隐藏层的层数，例如，一个具有 3 个隐藏层的神经网络被称为深度为 3 的神经网络。一般来说，深度学习模型的深度越大，可以学习到的抽象特征和复杂模式越多，模型的性能和泛化能力也就越好。在实践中，深度学习模型的深度可以根据具体任务和数据集的复杂度进行调整，以达到最佳性能。比如谷歌用来进行语音识别和图像搜索的 Google Brain 项目，所构建的人工神经网络有超过 10 亿个节点。

之前 AI 领域的实际应用主要是使用传统的机器学习算法，虽然这些传统的机器学习算法在很多领域都取得了不错的效果，不过仍然有非常大的提升空间。深度学习出现后，计算机视觉、

自然语言处理、语音识别等领域都取得了非常大的进步。

2009 年，辛顿教授与微软合作，将深度学习应用于语音识别中，结果在公开测试数据集上表现出色，成功将错误率降低了 30%。一石激起千层浪！这样大的提升给沉寂多年的语音识别领域带来了新的希望，因为在此之前，语音识别已经多年没有出现过什么显著的进展。此时辛顿已经 62 岁了。在此之前，他的工作成果一直属于边缘化的领域，他完全靠着信念在深度学习这个领域坚持着。

怎么才能让深度学习算法进化得更快一点呢？

公开竞赛是激励研究人员和工程师挑战极限的极好方法，让研究人员通过竞争来挑战共同基准，但难题是，数据从哪里来？

这要感谢一位华人科学家——李飞飞（如图 1-22 所示）。她通过互联网众包的方式，花了两年多时间，用大量人工对图像进行标记，建立了一个 1400 万张带标签图像的免费数据库——ImageNet，并于 2009 年上线。

图 1-22　李飞飞（来源 ted.com）

2010 年，斯坦福大学的李飞飞教授创办了一年一度的 ImageNet 挑战赛。这是一个图像识别比赛，需要基于 ImageNet 数据库，对多达 1000 个类别的图像做出分类，比赛的目的是鼓励计算机视觉方面的突破性进展。随着比赛不断举办，ImageNet 的名声越来越大，成为衡量图像识别算法性能如何的一个基准。

2012 年 10 月，辛顿教授带着两个学生亚力克斯 • 克里哲夫斯基（Alex Krizhevsky）和伊利亚 • 萨特斯基弗（Ilya Sutskever）参加 ImageNet 比赛（如图 1-23 所示），获得了冠军，他们的图片识别的正确率达到了 83.6% 的 top-5 精度，大大超过了第二名。

从此之后，深度学习方法包揽了这个比赛的冠军。到 2015 年，ImageNet 挑战赛获胜者的精度达到了 96.4%，也就是说，AI 分类的错误率比人工分类还低（如图 1-24 所示）。

图 1-23　萨特斯基弗（左）、克里哲夫斯基（中）和辛顿（右）（来源：多伦多大学网站）

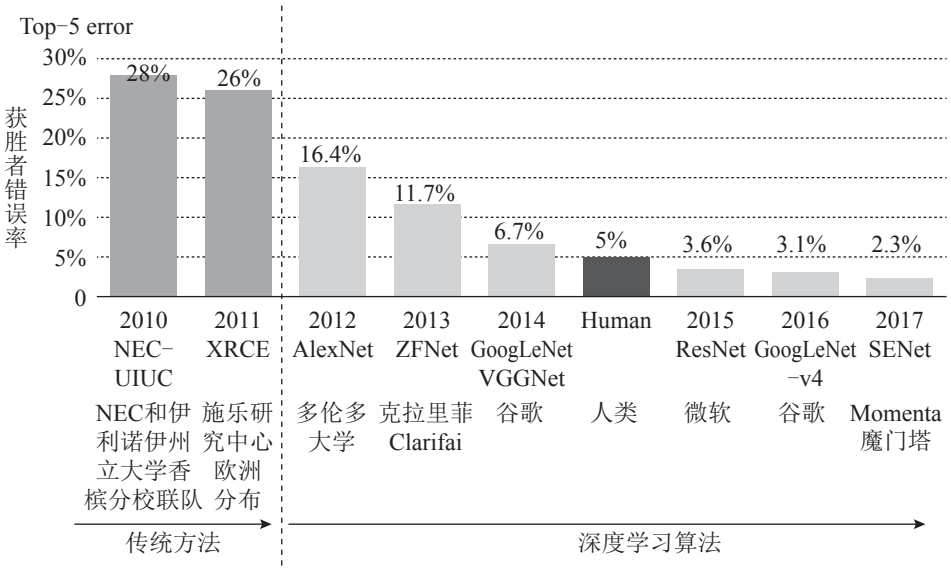


图 1-24　历届 ImageNet 挑战赛冠军（来源：论文《*Application of Deep Learning in Dentistry and Implantology*》）

直到今天，人们都把深度学习能够快速发展的原因归结于这场比赛，从那以后整个 AI 研究领域都发生了变化。

这次比赛不仅证明了深度学习的先进性，还显示了使用 GPU 加速功能的深度学习模型是多么强大。从此，GPU 赢得了它的狂热地位和主流媒体的关注，引发了深度学习革命。

在辛顿和他的两名学生拿到冠军，并且发表论文介绍了算法 AlexNet 后，有一个中国人敏锐地看到了机会，他就是曾在 2010 年带领美国 NEC 实验室拿过第一届 ImageNet 竞赛冠军的余凯。余凯深知这篇论文背后的重要意义——神经网络技术的突破，之后超 6 万次的引用量也确实证明了 AlexNet 的论文是计算机科学史上最有影响力的论文之一。

当时余凯已经离开硅谷的 NEC 实验室，回到北京，加入了百度，领导百度新成立的多媒体部，这个部门包括了语音识别团队和图像识别团队。

余凯立刻写了封电子邮件给辛顿，表达了百度要和他深入合作的想法。

很快，辛顿回复了愿意合作，并且提出了 100 万美元科研经费的需求。在与百度 CEO 李彦宏沟通并获得支持后，余凯爽快地答应了辛顿的合作条件，并给出了 1200 万美元的报价。

没想到的是，如此爽快答应给出巨额资金让辛顿意识到了巨大的机会。

辛顿问律师，如何让他的新公司具有最大的价值，尽管当前只有三名员工，既没有产品，也没有底蕴。律师给他的选择之一是设立一个拍卖会。

辛顿马上注册了一家公司——DNNresearch，公司只有三个人：辛顿与他的两个研究生学生亚力克斯·克里哲夫斯基和伊利亚·萨特斯基弗，也就是 ImageNet 比赛的参赛队伍，这家公司隶属于多伦多大学计算机科学院。

然后，辛顿又找来了谷歌、微软以及当时还名不见经传的 DeepMind，在太浩湖以秘密竞拍的方式做团队收购。拍卖以电子邮件形式举行，四家竞拍者身份互相保密。最终，谷歌报价 4400 万美元力压百度赢得拍卖。

这次拍卖，也意味着辛顿 30 多年的耕耘终于开始得到回报。

虽然辛顿在人工神经网络领域早就是一位泰斗级人物，1998 年就被选为英国皇家学会院士，硕果累累，荣誉等身，但他在学术上的成就，还是抵不过大众脑海里“神经网络没有前途”的偏见。在很长一段时间里，多伦多大学计算机系私下流行着一句对新生的警告：“不要去辛顿的实验室，那是没有前途的地方。”即便如此，辛顿依然不为所动，仍坚持自己的神经网络研究方向没有丝毫动摇。在这 30 多年的时间里，神经网络相关学术论文都很难得到发表，但辛顿仍坚持写了 200 多篇不同方法和方向的研究论文，这为后来神经网络的多点突破打下了坚实基础。

对于谷歌来说，收购 DNNresearch 只是一个开始，2013 年谷歌又宣布以 5 亿美元收购了 DeepMind。

谷歌不断收购深度学习领域的公司，其最主要目的是抢到一批世界上的一流专家。在一个迅速成长的 AI 领域里面，顶尖专家能带来突破性成果，拉开与其他对手的差距。

不要觉得辛顿的研究生学生不算顶尖专家，这里提一下伊尔亚·苏茨克维（Ilya Sutskever，如图 1-25 所示）。

在多伦多大学读本科时，苏茨克维想加入辛顿教授的深度学习实验室。一天他直接敲开了辛顿教授办公室的门，询问自己是否可以加入实验室。辛顿让他提前预约，但苏茨克维不想再浪费时间，他问：“就现在怎么样？”

辛顿意识到苏茨克维是一个敏锐的学生，于是给了他两篇论文让他阅读。一周后，苏茨克维回到教授办公室，然后告诉辛顿他不理解。

“为什么不理解？”辛顿问。

图 1-25　伊尔亚·苏茨克维（来源：多伦多大学网站）

苏茨克维解释说："人们训练神经网络来解决问题，当人们想解决不同问题时，就得用另外的神经网络重新开始训练，但我认为人们应该有一个能够解决所有问题的神经网络。"

这个回答显示了苏茨克维的独特思考能力，辛顿非常赞赏，于是向他发出邀请，让他加入自己的实验室。

在 DNNresearch 被谷歌收购后，苏茨克维进入谷歌工作，他发明了一种神经网络的变体，能将英语翻译成法语。他提出了"序列到序列学习"（Sequence to Sequence Learning），它能捕捉到输入的序列结构（如英语的句子），并将其映射到同样具有序列结构的输出（如法语的句子）。

他说，研究人员本不相信神经网络可以做翻译，所以当它们真的能翻译时，这就是一个很大的惊喜。他的发明比起之前的机器翻译，在错误率上大大减少，还让谷歌翻译实现了跨越式的大升级。

神经网络翻译是怎么做到的呢？传统的统计机器翻译通常以英语为主要语言，因此在将俄语翻译成德语时，机器必须先将文本翻译成英语，再将英语翻译成德语。这样做会造成双重的信息损失。相比之下，神经网络翻译则不需要经过翻译为英语这一步骤，它只需要一个解码器，就能够实现在没有词典可查时进行不同语言之间的翻译。换句话说，即使没有使用词典，神经网络仍然能够进行翻译。它通过分析大量的语言数据来学习翻译的规则和模式，实现了自动翻译。

自从使用了神经网络进行翻译，单词顺序错误率降低了 50%，词汇错误率降低了 17%，语法错误率降低了 19%。神经网络甚至学会了用不同的语言来调整词性与大小写。

如今，谷歌翻译已经实现了 103 种语言的翻译，覆盖了地球上 99% 的人口。

在谷歌工作两年后，苏茨克维离开了那里，进入一家非营利机构，薪资比谷歌低不少。这家机构就是 OpenAI，苏茨克维作为联合创始人和首席科学家，在 ChatGPT 的开发中居功至伟。

3. **深度学习**（2012 年至今）

从 2012 年开始，在强大算力的支持下，在深度学习算法的带领之下，AI 迎来了第三次浪潮，也就是今天我们正在经历的这股 AI 浪潮。

深度卷积神经网络（ConvNet）已成为所有计算机视觉任务的首选算法，并在许多其他类型的问题上也得到了应用，比如棋类比赛、自然语言识别等。大批研究资金涌入这一领域，大公司广泛采用这一算法——脸书用它来标记用户照片；特斯拉自动驾驶汽车用它来检测物体。

深度学习发展得如此迅速，主要原因在于它在很多问题上都表现出很好的性能，但这并不是唯一的原因。深度学习还让解决问题变得简单，因为它将特征工程完全自动化，而这曾经是机器学习工作流程中最关键的一步。

浅层学习的机器学习技术通常采用简单的变换，无法准确表达复杂问题，因此需要进行大量的特征工程，手动设计数据表示层，以适应这些方法的处理。深度学习技术将这一步骤完全自动化了，一次性学习所有特征，无须手动设计，大大简化了机器学习的工作流程。在实际应用中，这种方法省时省力，也省去了人工成本。

在短短三年里，数十家创业公司涉足深度学习领域，从业人数由几千人增加到数万人。

其中值得一提的有一家中国公司——商汤科技。

2014 年，香港中文大学教授汤晓鸥（如图 1-26 所示）领导的计算机视觉研究组开发了名为 DeepID 的卷积神经网络深度学习模型。该模型将每张输入的人脸表示为 160 维向量，并通过其他模型进行分类。在 LFW（Labeled Faces in the Wild，人脸识别领域的测试基准）数据库上，该人脸识别技术的识别率为 99.15%，而人类肉眼在 LFW 上的识别率为 97.52%。这表明，深度学习已经在学术研究层面上超越了人类肉眼的识别能力，这标志着人脸识别技术进入了一个新时代。

同年，汤晓鸥带领实验室的成员正式创办了商汤科技。2021 年 12 月 30 日，商汤科技在香港证券交易所上市，在 IPO 后的四个交易日内，商汤科技股价涨幅超过 130%，总市值达到 2074 亿港元，成为 AI 领域全球最大 IPO。

图 1-26　汤晓鸥（来源：ustcif.org.cn）

（1）生成对抗网络 GAN

2014 年，伊恩・古德费洛（Ian Goodfellow, 如图 1-27 所示）及本吉奥等人提出了生成对抗网络 GAN（Generative Adversarial Network，如图 1-28 所示）。GAN 被誉为近年来最酷炫的神经网络，是无监督学习最具前景的方法之一。

图 1-27　伊恩・古德费洛（来源：dukakis.org）

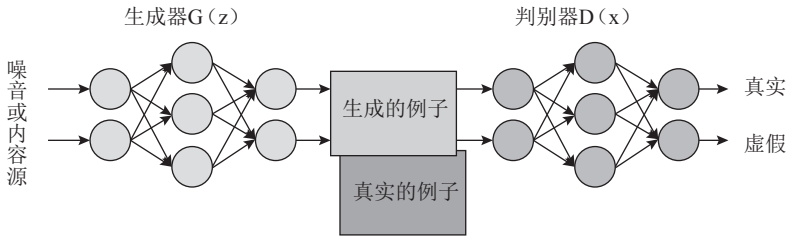


图 1-28　GAN 工作原理

简而言之，GAN 是一种“道高一尺，魔高一丈”的博弈算法，正如同其名字“对抗生成”，它使用了两个神经网络：一个作为生成器（Denerator），另一个作为判别器（Discriminator）。

在训练过程中，生成器的目标是生成越来越好的样本去使得判别器失效，而判别器则是要提升自己的判断能力使自己不被骗。在这样训练的博弈过程中，分别提高了两个模型的生成能力和判别能力。

作为图像分析应用中的里程碑式成果，GAN 能将图像放大到超分辨率，通过图像语义分析，使模糊不清的图像文件变得清晰可辨。后来谷歌推出的知名的 Auto Draw（如图 1-29 所示）就是基于这项技术。

图 1-29　谷歌 Auto Drew 技术

GAN 模型一问世就风靡人工智能学术界，在多个领域得到了广泛应用。它也随即成为很多 AI 绘画模型的基础框架，其中生成器用来生成图片，而判别器用来判断图片质量。

虽然 GAN 在图像生成应用上最为突出，但在计算机视觉中还有许多其他应用，如图像绘画、图像标注、物体检测和语义分割。在自然语言处理中应用 GAN 的研究也呈现一种增长趋势，如文本建模、对话生成、问答和机器翻译。

（2）TensorFlow 框架

2015 年，谷歌开源了 TensorFlow 框架。这是一套综合性的机器学习系统框架，是一个基于数据流编程（dataflow programming）的符号数学系统，被广泛应用于各类机器学习（machine learning）算法的编程实现，它可以很好地支持深度学习的各种算法，并支持多种计算平台，系统稳定性较高。

它的开发者是谷歌内部最早系统性地研究 AI 技术的团队——Google Brain。TensorFlow 是谷歌的第二代机器学习工具，其前身是谷歌的神经网络算法库 DistBelief。

在开源之前，TensorFlow 仅供谷歌公司内部使用。据说谷歌内部有超过 4000 个项目都能找到 TensorFlow 的配置文件，几乎所有团队都在采用此项技术，包括搜索排名、应用商城推荐、Gmail 反垃圾邮件以及 Android 系统等产品团队。

谷歌希望 TensorFlow 这款开源的 AI 工具包能像当年的安卓一样赋能全世界的开发者们，所以它也被称为“AI 界的安卓系统”。

例如，美国航天总署有个开普勒计划，目标是通过望远镜持续不断地观察太空中恒星亮度的变化，希望发现太阳系以外的行星系统，最终希望发现另外一个适宜人类居住的行星。目前该计划已经积累了上百亿个观察数据，他们使用了 TensorFlow 的模型，帮助科学家发现了 2500 光年以外的开普勒 90 星系中的第八颗行星。

也有科学家利用 TensorFlow 把语音处理技术用到鸟类保护上。在丛林里安装很多收音器来采集鸟类的声音，通过 TensorFlow 模型对其进行分析，就可以很准确地估算出鸟类在一片森林中的数量，从而可以更加精准地实施保护。

TensorFlow 的优势是既可以在智能手机端应用，也可以在大规模图形处理单元集群上运行。换句话说，TensorFlow 可以进行异构设备的分布式计算，它能够自动适应各种平台，并在不同设备上运行模型，无论是手机、单个 CPU/GPU，还是成百上千个 GPU 卡组成的分布式系统。

TensorFlow 的开源，大大降低了深度学习在各个行业的应用难度，如语音识别、自然语言理解、计算机视觉和广告等。

2015 年末，谷歌首席执行官桑达尔·皮查伊（Sundar Pichai）表示：“机器学习这一具有变革意义的核心技术将促使我们重新思考做所有事情的方式。我们用心将其应用于所有产品，无论是搜索、广告、YouTube 还是 Google Play。我们尚处于早期阶段，但你将会看到我们系统性地将机器学习应用于这些领域。”

由于 TensorFlow 这样深度学习工具集的大众化，任何人只要具有基本的 Python 脚本技能，

就可以从事高级的深度学习研究，TensorFlow 很快就成为大量创业公司和研究人员转向该领域的首选深度学习解决方案，这又进一步推动了深度学习的进展。

（3）AlphaGo

在 AlphaGo 横空出世之前，几乎没有人相信 AI 可以在围棋上胜过人类。因为围棋的复杂度太高，棋盘有 361 个交叉点，每个点有三种状态，也就是说共有 3 的 361 次方变化的可能，即围棋的着数变化是 10 的 172 次方，这比太阳系里所有的原子数量还多得多。

但在 2016 年 3 月，AlphaGo 赢了围棋顶尖高手李世石（如图 1-30 所示）。

图 1-30　AlphaGo 和李世石对战（来源：time.com）

因为出现了以卷积神经网络为代表的深度学习技术，让 AI 有了飞跃发展。AlphaGo 使用了多层卷积神经网络来分析和理解棋盘上的局面，从而预测出最可能的下棋位置。

2017 年 5 月 23 日，升级版 AlphaGo 出场，与当时世界排名第一的围棋棋手柯洁进行了对弈（如图 1-31 所示）。这个版本让三子还能胜过之前战胜过李世石的 AlphaGo 版本。

图 1-31　柯洁对战 AlphaGo（来源：tech-camp.in）

在赛前，柯洁发微博感叹：“早就听说新版 AlphaGo 的强大，但……让……让三子？我的天。这个差距有多大呢？简单地解释一下就是一人一手轮流下的围棋，对手连续让你下三步，又像武林高手对决让你先捅三刀一样。我到底是在和一个怎样可怕的对手下棋……”

不出所料，新版 AlphaGo 以 3：0 的战绩赢了柯洁。并不是柯洁没发挥好，其实早在 4 个多月前，这个新版的 AlphaGo 以“Master”为名，在网上挑战中韩日顶尖高手，60 战全胜。

AI 是怎么做到的呢？

没人知道确切的细节，AI 就像一个黑盒子，吞掉一大堆学习材料之后突然说：“我会了。”一测试，你发现它真的会了，可是你不知道它掌握的究竟是什么，因为神经网络本质上只是一大堆参数，无法被解释。

虽然会下棋的 AI 没有很高的商业价值，但它带动了 AI 的快速发展。DeepMind 团队又在同样的技术基础上开发了 AlphaFold，可以通过计算预测蛋白质结构，对生物研究和制药都有巨大的推动。

AlphaGo 的成功也将强化学习推向了新高潮。在 AlphaGo 的开发中，DeepMind 团队使用了无监督学习技术来让 AlphaGo 从大量的未标注围棋数据中自动学习，提取围棋游戏中的模式和结构。

具体来说，DeepMind 团队将 AlphaGo 暴露在大量的未标注围棋数据中，例如棋谱、游戏记录和围棋书籍等。通过这些数据，AlphaGo 逐渐学会了如何下围棋，如何识别游戏中的模式和结构，以及如何根据这些模式和结构来制定自己的下棋策略。

什么是无监督学习？

所谓无监督学习（unsupervised learning），是指不用标记每个数据是什么，AI 看多了会自动发现其中的规律和联系。能够无监督学习的 AI 就好像一个聪明的学者，自己从资料里面学习，看了大量的内容，看多了就会了。

无监督学习是怎么知道对错的呢？

以语言模型为例，最简单的一种做法，就是拿掉句子的一个词，然后让 AI 猜测是哪一个词，因为有原句作为标准答案，就可以训练 AI 来猜测到正确的结果。谷歌的 BERT 模型就是类似这样的训练机制，训练后的模型在阅读理解方面的成绩非常好。

1.4　预训练大模型时代

前面讲到了符号主义、连接主义和行为主义三个流派，它们好像各不相干，但实际上，这些不同的理论可以互相补充。在 AI 技术不断发展和应用时，AI 系统往往会用到多个流派的理论和技术，既有神经网络，又有强化学习，还有推理规则等，因为机器学习单项发展有其局限性，通过各种流派技术的组合才能实现更加复杂和强大的功能。

ChatGPT 就是成功运用了行为主义的强化学习技术，采用无监督学习的算法来训练，完成海量的语料学习。

现代的 AI 以大模型为主，集成的技术很多，已经很难简单判定是属于哪一个流派了，如 ChatGPT 这样的大模型，就使用了以上三个流派的技术。

1.4.1 预训练大模型

为什么需要预训练大模型呢？

因为一个未经训练的 AI 模型并不能直接用。

这就好像人的大脑，虽然有很多的神经网络结构，但如果没学习过，就是个文盲。同理，大模型在训练之前，只有搭建好的网络结构和几万甚至几千亿个参数，需要把大量的素材“喂”给它进行训练。每个素材进来，神经网络训练一遍后，各个参数的权重就会进行相关的调整，这个过程就是机器学习。等训练得差不多了，就可以把所有参数都固定下来，预训练模型就“炼制”完成了。

除了训练，推理也是必要的。在 ChatGPT 模型中，推理是通过预训练的 Transformer 模型实现的。该模型通过不断试验和迭代，自我学习并调整参数，提高了推理能力，在实践中可以表现出更好的结果。符号主义的方法可用于设计基于规则的框架，以处理和表达基于符号和关系的知识，从而使 AI 模型具备推理或逻辑推断的能力。

从 2016 年开始，出现了越来越大的 AI 模型，模型参数规模越大，性能表现也越好。模型的预训练也起到了很关键的作用，这个以大模型为主的发展路径跟之前的 AI 相比又有了很大的变化。

2022 年，基于 AI 模型，出现了许多 AIGC 产品，尤其是在 AI 绘画方面，典型的有 DALL-E 2、Midjourney、Stable Diffusion 和 Disco Diffusion 等。

下面以谷歌大语言模型的简史来讲述 AI 中的关键技术发展。

1.4.2 2017 年 6 月，6500 万个参数的 Transformer

2017 年 6 月，谷歌大脑团队（Google Brain）在神经信息处理系统大会（NeurIPS，该会议为机器学习与 AI 领域的顶级学术会议）发表了一篇名为《自我注意力是你所需要的全部》（*Attention Is All You Need*）的论文。作者在文中首次提出了基于自我注意力机制（self-attention）的变换器（transformer）模型，并首次将其用于理解人类的语言，即自然语言处理。

在这篇文章发布之前，自然语言处理领域的主流模型是循环神经网络（RNN，recurrent neural network）。循环神经网络模型的优点是能更好地处理有先后顺序的数据，它被广泛用于自然语言处理中的语音识别、手写识别、时间序列分析以及机器翻译等。但这种模型也有不少缺点：在处理较长序列，例如长文章、书籍时，存在模型不稳定或者模型过早停止有效训练的问题，以及训练模型时间过长的问题。

从 NLP 发展的逻辑来看，最早的 NLP 模型是基于对单个单词统计来做的；到后来卷积神经网络（CNN）出现，机器开始能够基于两三个单词来理解词义；再往下发展到 RNN 时代，这时 AI 基本上就可以沿着整个序列（sequence）进行积累，可以理解相对长的短语和句子，不过依然无法真正理解上下文。

论文中提出的 Transformer 模型（如图 1-32 所示）中提出的注意力机制（attention model）是一个很重要的突破。在这个阶段，AI 开始能够结合所有上下文，理解各个词表达的重要性不同。这就很像我们的快速阅读，为什么人类能够做到“一目十行”，是因为我们能看到一些关键词，而每个词的重要性不一样。通过快速扫描方式，找到关键的内容就搞懂了内容，并不需要在每个词上花同样的注意力。

“注意力机制”正是起到了这个作用，它告诉 AI 各个关键词之间的关系如何，谁重要和谁不重要。

Transformer 能够同时进行数据计算和模型训练，训练时长更短，并且训练得出的模型可用语法解释，也就是模型具有可解释性。

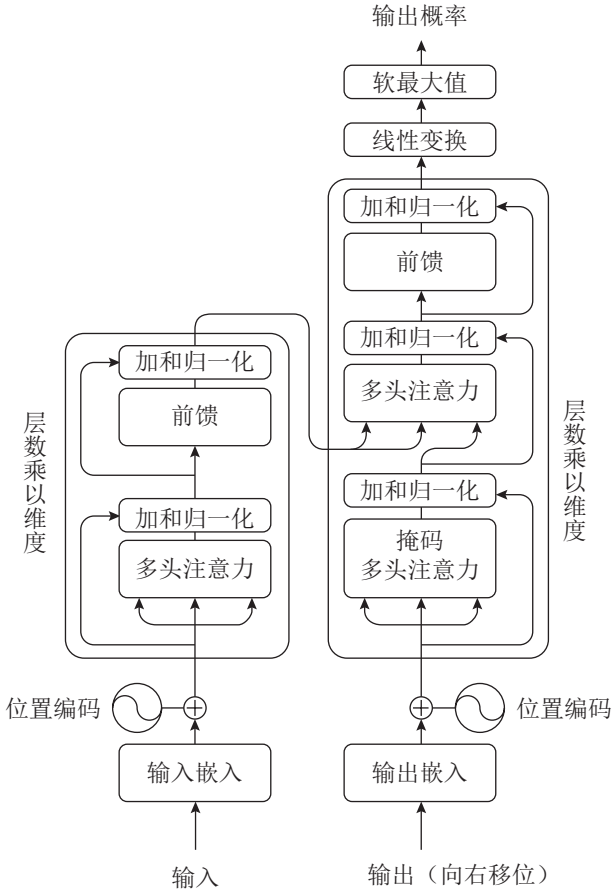


图 1-32　最初变换器模型的架构（来源：论文《*Attention Is All You Need*》）

谷歌大脑团队使用了多种公开的语言数据集来训练最初的 Transformer 模型，一共有 6500 万个可调参数。

经过训练后，这个最初的 Transformer 模型（如图 1-33 所示）在翻译准确度、英语成分

句法分析等各项评分上都达到了业内第一，成为当时最先进的大型语言模型（Large Language Model, LLM），其最常见的使用场景就是输入法和机器翻译。

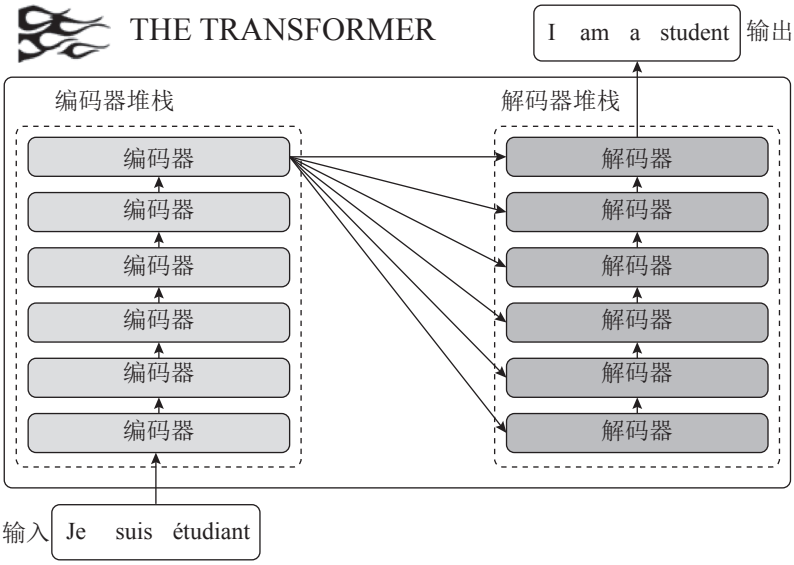


图 1-33　图解 Transformer 翻译（来源：jalammar.github.io）

Transformer 模型自诞生的那一刻起，就深刻地影响了接下来几年 AI 领域的发展轨迹。

因为谷歌大脑团队在论文中提供了模型的架构，任何人都可以用其来搭建类似架构的模型，并结合自己手上的数据进行训练。

于是 Transformer 就像其另一个霸气的名字“变形金刚”一样，被更多人研究，并不断变化。

短短几年里，该模型的影响已经遍布 AI 的各个领域——从各种各样的自然语言模型到预测蛋白质结构的 AlphaFold2 模型，用的都是它。

1.4.3　2018 年 10 月，3.36 亿个参数的 BERT

2018 年 10 月，谷歌提出 3.36 亿个参数的 BERT（Bidirectional Encoder Representation from Transformers，来自 Transformers 的双向编码表示）模型。

BERT 的特点是可以使用大量没有标注的数据，自己创建一些简单任务来进行自学。怎么学呢？比如一句话，AI 会把其中的一个词藏起来，然后猜这个词应该是什么，有点像机器自己和自己玩游戏，这样它的语言理解能力就会变得越来越强。

在 SQuAD1.1 机器阅读理解顶尖水平测试中，BERT 表现惊人：两个衡量指标全部超越人类，且在 11 种不同的 NLP 测试中获得最佳成绩，例如将 GLUE 基准提升至 80.4%（绝对改进 7.6%），以及 MultiNLI 准确度达到 86.7%（绝对改进 5.6%），成为 NLP 发展史上里程碑式的模

型成就。

据测试，在同等参数规模下，BERT 的效果好于 GPT-1。因为它是双向模型，可以同时利用上文和下文来分析，而 GPT-1 是单向模型，无法利用上下文信息，只能利用上文（如图 1-34 所示）。

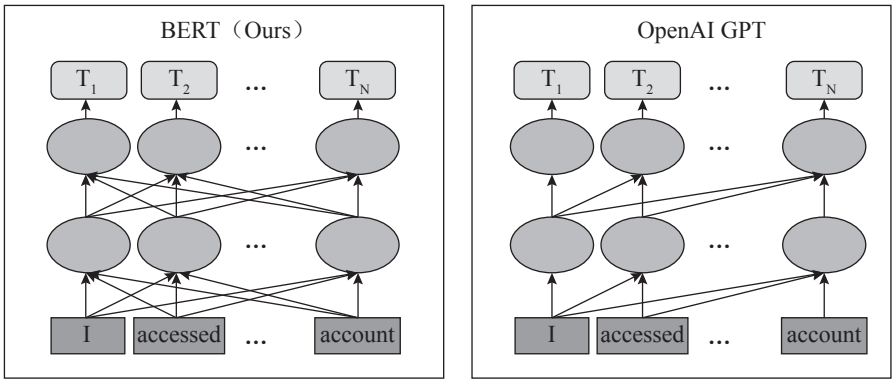


图 1-34　BERT 和 GPT-1（来源：hackernoon.com）

GPT-1 学会猜测句子中的下一组单词，而 BERT 则学会猜测句子中任何地方缺少的单词。这有很多好处：如果给 BERT 几千个问题和答案，它可以自己学会回答其他类似的问题。另外，BERT 也可以学会如何进行对话。

从阅读理解方面来看，BERT 模型的提升是很大的。在当时的 SQuAD 竞赛排行榜上，排在前列的都是 BERT 模型，阅读理解领域也基本上被 BERT 霸榜了。谷歌的 BERT 模型完胜当时所有的大模型。

1.4.4　2019 年 10 月，110 亿个参数的 T5

2019 年 10 月，谷歌在论文《探索使用统一文本到文本变换器的转移学习极限》（*Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*）中提出了一个新的预训练模型：T5（如图 1-35 所示）。该模型涵盖了问题解答、文本分类等方面，参数量达到了 110 亿个，成为全新的 NLP SOTA 预训练模型。在 SuperGLUE 基准上，T5 也超越了 Facebook 提出的 RoBERTa，以 89.8 的得分成为仅次于人类基准的顶尖 SOTA 模型。

为什么叫 T5？因为这是“Text-To-Text Transfer Transformer”的缩写（五个 T）。

T5 作为一个文本到文本的统一框架，可以将同一模型、目标、训练流程和解码过程，直接应用于实验中的每一项任务。研究者可以在这个框架上比较不同迁移学习目标、未标注数据集或者其他因素的有效性，也可以通过扩展模型和数据集来发现 NLP 领域迁移学习的局限。

2022 年 6 月发布的 Flan-T5 语言模型，通过在超大规模的任务上对 T5 进行微调，使它具备极强的泛化性能，在 1800 多个不同的 NLP 任务上都有很好的表现。

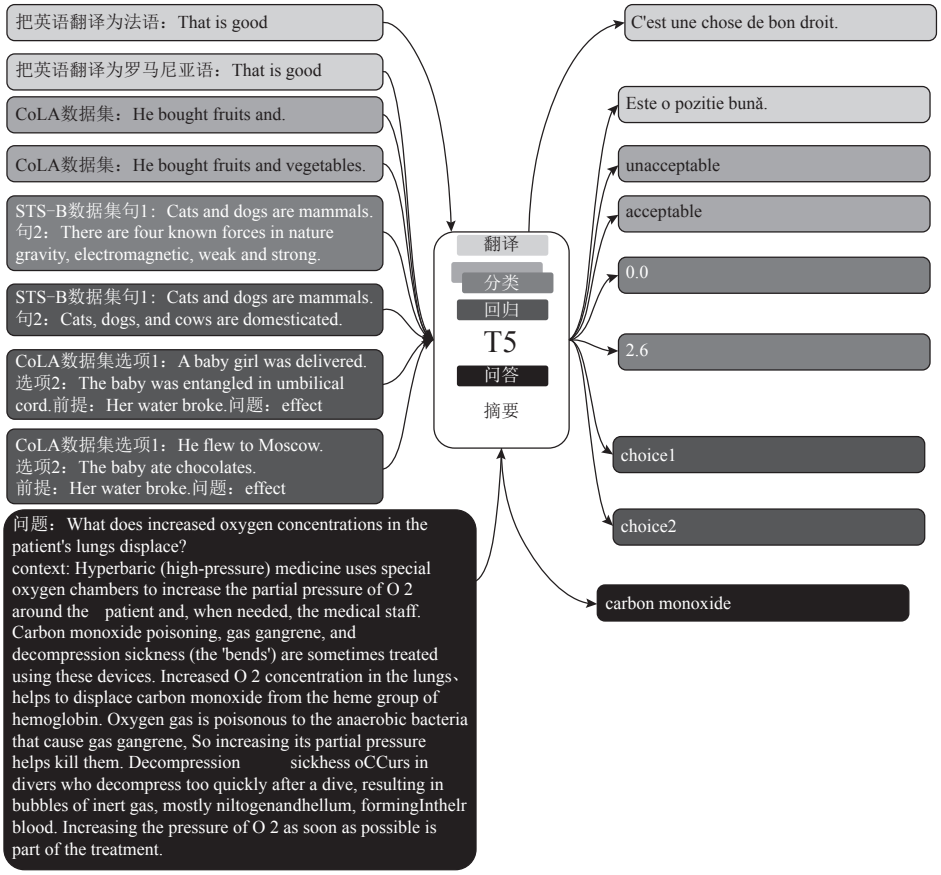


图 1-35　T5 模型示意图（来源：towardsdatascience.com）

微调的目的是让语言模型理解指令并学会泛化，而不仅是学会解决给定的任务。这样，当模型面对现实世界的新任务时，只需学习新的指令就能解决任务。一旦模型训练完毕，就可以在几乎全部的 NLP 任务上直接使用，实现一个模型解决所有问题（One model for all tasks），这就非常有诱惑力！

从创新来看，T5 算不上出奇制胜，因为模型没有用到什么新方法，而只是从全面的视角来概述当前 NLP 领域迁移学习的发展现状。它的成功，是通过“大力出奇迹”，用 110 亿个参数的大模型，在摘要生成、问答和文本分类等诸多基准测试中都取得了不错的性能，一举超越现有最强模型。

谷歌编写的 T5 通用知识训练语料库中的片段来自 Common Crawl 网站，该项目每个月从网络上爬取大约 20TB 的英文文本。

具体做法分为三步：

（1）任务收集：收集一系列监督的数据，这里一个任务可以被定义成“数据集，任务类型的形式”，比如“基于 SQuAD 数据集的问题生成任务”。

（2）形式改写：因为需要用单个语言模型来完成超过 1800 多种不同的任务，所以需要将任务都转换成相同的“输入格式”给模型进行训练，同时这些任务的输出也需要是统一的“输出格式”。

（3）训练过程：使用恒定的学习速率和 Adafactor 优化器进行训练；同时将多个训练样本打包成一个训练样本，这些训练样本通过一个特殊的“结束标记”进行分割。在每个指定的步数，进行“保留任务”上的模型评估，并保存最佳的检查点。

尽管微调的任务数量很大，但相比语言模型本身的预训练过程，计算量小了很多，只有 0.2%。所以通过这个方案，大公司训练好的语言模型可以被再次有效地利用，应用方只需要做好“微调”即可，不用重复耗费大量计算资源去训练一个语言模型。

从竞赛排行榜看，T5 以绝对优势胜出。

1.4.5　2021 年 1 月，1.6 万亿个参数的 Switch Transformer

2021 年 1 月，在 OpenAI 的 GPT-3 发布仅几个月后，谷歌大脑团队就重磅推出了超级语言模型 Switch Transformer，它有 1.6 万亿个参数，是 GPT-3 参数量的 9 倍。看起来，大模型的“大”成为竞争的关键。

研究人员在论文中指出，大规模训练是通向强大模型的有效途径，具有大量数据集和参数计数的简单架构可以远远超越复杂的算法，但目前有效的大规模训练主要使用稠密模型，而 Switch Transformer 采用了“稀疏激活”技术。所谓稀疏，指的是对于不同的输入，只激活神经网络权重的子集。

根据作者介绍，Switch Transformer 是在 MoE 的基础上发展而来的。MoE 是 20 世纪 90 年代初首次提出的 AI 模型，它将多个“专家”或专门从事不同任务的模型放在一个较大的模型中，并有一个“门控网络”来选择对于任何给定数据要咨询哪些 / 个“专家”。尽管 MoE 取得了一些显著成果，但复杂性、通信成本和训练不稳定阻碍了其被广泛采用。

Switch Transformer 的新颖之处在于，它有效地利用了为稠密矩阵乘法（广泛用于语言模型的数学运算）而设计的硬件——例如 GPU 和 Google TPU。研究人员为不同设备上的模型分配了唯一的权重（如图 1-36 所示），因此权重会随着设备的增多而增加，但每个设备上仅有一份内存管理和计算脚本。

Switch Transformer 在许多下游任务上有所提升。研究人员表示，它可以在使用相同计算资源的情况下使预训练速度提高 7 倍以上。他们证明，大型稀疏模型同样可以用于创建较小的、稠密的模型，并且通过微调，这些模型相对大型模型会有 30% 的质量提升。

Switch Transformer 模型在 100 多种不同语言之间的翻译测试中，研究人员观察到了“普遍改进”，与基准模型相比，91% 的语言翻译有 4 倍以上的提速。

研究人员认为，在未来的工作中，Switch Transformer 可以应用到其他模态或者跨模态的研究当中。模型稀疏性可以在多模态模型中发挥出更大的优势。

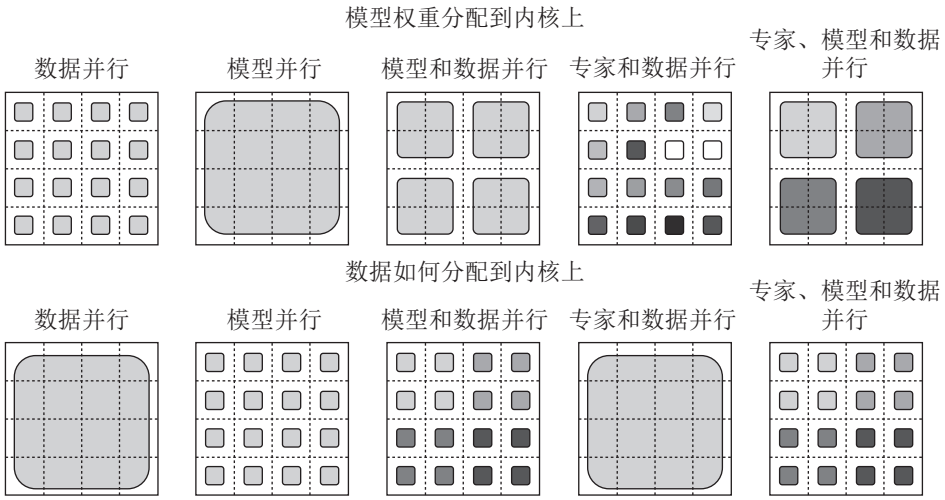


图 1-36　模型权重分布示意图（来源：towardsdatascience.com）

从结果来看，这个模型意味着谷歌的新模型在翻译等领域获得了绝对的胜利。

但从另一方面看，模型越大，部署的难度越大，成本也越高，所以从产出结果的效率来看是低的，这就意味着它未必能赢得最终的胜利。

这也便能解释为什么 Switch Transformer 这样开源的万亿参数模型影响力不大，许多人根本没听说过。

1.4.6　2021 年 5 月，1370 亿个参数的 LaMDA

2021 年 5 月的 Google I/O 大会上，谷歌展示了其最新的 AI 系统 LaMDA（Language Model for Dialogue Applications，对话应用语言模型），它拥有 1370 亿个参数，略少于 GPT-3，但比 13 亿个参数的 InstructGPT 多 100 多倍。

不过，LaMDA 跟其他语言模型都不同，因为它专注于生成对话，跟 ChatGPT 一样，LaMDA 可以使回答更加“合情合理”，让对话更自然地进行，其目的不是提供信息搜索，而是通过对自然语言问题的回答来帮助用户解决问题（如图 1-37 所示）。但跟 ChatGPT 不一样的是，它可以利用外部知识源展开对话。

而且，这些回复都不是预先设定的，与它进行多次对话时，同一个问题不会出现相同的答案。

当时，这个演示让技术圈轰动了。

这么牛的对话机器人，按说应该像 ChatGPT 这样迅速火爆才是。

实际上，没有多少人了解 LaMDA。

因为直到 2023 年 2 月，谷歌才向公众发布了 LaMDA 支持的 Bard 对话机器人。部分原因在于 LaMDA 存在较高的误差，且容易对用户造成伤害，此类瑕疵被谷歌称之为有“毒性”。

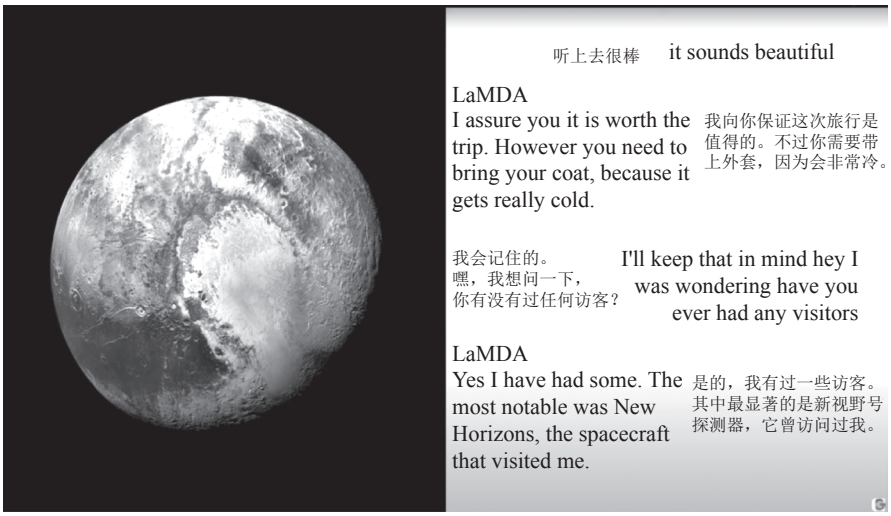


图 1-37　对话过程（来源：Google I/O 大会视频）

谷歌的 CEO 桑达尔·皮查伊和谷歌 AI 部门长期负责人 Jeff Dean 表示，谷歌其实完全有能力拿出类似 ChatGPT 的成果。只是一旦出了纰漏，谷歌这样的企业巨头无疑需要承担更高的经济和声誉成本。

因为全球有数十亿用户在使用谷歌的搜索引擎，而 ChatGPT 到 12 月初才刚刚突破 100 万名用户。

那么，在这一局，虽然谷歌最早发布了产品，看起来有不错的结果，毕竟能采用外部知识的对话机器人更有时效性价值，但遗憾的是，谷歌交卷太晚了。而且从使用的千亿个参数来看，这个语言模型的效率比不上 OpenAI 的 InstuctGPT 语言模型。

1.4.7　2022 年 4 月，5400 亿个参数的 PaLM

2022 年 4 月，谷歌首次发布了 PaLM（Pathways Language Model）大语言模型，使用了 5400 亿个参数进行训练，约是 GPT-3 参数量的三倍。PaLM 是一个只有解码器的密集 Transformer 模型。这个模型训练使用了 6144 块 TPU。

PaLM 使用英语和多语言数据集进行训练，包括高质量的 web 文档、书籍、维基百科、对话和 GitHub 代码。PaLM 在许多非常困难的任务上显示出了突破性的能力，包括语言理解、生成、推理和代码等相关任务。

2023 年 5 月，谷歌在 Google I/O 大会上发布了升级版 PaLM 2，宣布 PaLM 2 已被用于支持谷歌自家的 25 项功能和产品，其中包括 AI 聊天机器人 Bard、Gmail、谷歌 Docs、谷歌 Sheets 和 YouTube 等。

PaLM 2 是谷歌最先进的大语言模型，它擅长数学、编码、推理、多语言翻译和自然语言生成。它现在可以理解 100 多种语言，在高级语言能力考试中能达到“精通”的水平。

PaLM 2 也是多模态的，对标 GPT-4，可以同时处理文本和图像等不同类型的数据。用户可以发送一张厨房货架上食材的图片，并询问可以做什么菜。Bard 就会根据从图片中识别出来的食材给出一个合适的菜谱，并附上图片和步骤。除了图片，Bard 还可以处理音频、视频等其他类型的数据，并给出相应的回应。

PaLM 2 分为四种规格，从小到大依次为壁虎（Gecko）、水獭（Otter）、野牛（Bison）和独角兽（Unicorn），它们分别依据特定领域的数据进行了微调，以便为企业客户执行某些任务。其中最轻量级的“壁虎”版本能在移动设备上快速运行，离线状态下每秒可处理 20 个标记（token）。

为什么会有这么“小”的大模型？

因为在实际落地中，大模型不是参数量越大越好，在一些数据量小、任务并不复杂的场景中，追求泛化能力强但规模庞大的大模型，无异于“大炮打蚊子”，这时将大模型核心的泛化能力快速适配至不同场景才是关键。提供不同规模的 PaLM 2 意味着其落地应用会更加方便，可以面向不同的客户，部署在不同企业环境中，用户能够直接拿来用。

用上了 PaLM 2 的谷歌搜索，不仅对用户搜索语言的理解力更强，而且能把搜索结果的内容进行总结。在搜索结果的最上层显示为“AI Snapshot”（AI 快照）部分。

此前发布的 Bard 是基于 LaMDA 开发的，现在转用了 PaLM 2，其生产内容的能力得到了很大的提升。而且，谷歌将 Bard 跟旗下和外部产品的集成，让 Bard 可以打开谷歌地图、图片、视频、外部链接等多元化的信息，用户还可以将这些问题及答案一键导出到 Gmail、谷歌文档和表格之中。

Bard 的代码功能也很强，开发者们可以把 Bard 生成的代码进行导出，不仅能将其发送到谷歌的 Colab 平台，还能和另一个基于浏览器的 IDE Replit 一起使用。

在 ChatGPT 出来之前，谷歌虽然在 AI 大模型上投入巨大，研发了多个大模型，但由于各自为政，都没有达到量变到质变的临界点。没有一个大语言模型能与 ChatGPT 抗衡，更没有带来 AI 技术在应用方面的规模化突破。

而谷歌最新发布的 PaLM 2 大模型则是直接对标 GPT-4，用“大力出奇迹”，全面赋能谷歌生态的产品，交出了一份不错的答卷。

1.4.8 总结

预训练模型的价值在于用大量语料训练出一个懂很多知识的模型出来。OpenAI 在微软的支持下，投入巨大的算力训练 GPT-3 之后，又把 GPT-3 应用到了绘画、编程等领域，而程序代码方面的训练，对 GPT 的能力有一个很大的提升。谷歌的 PaLM 2 大模型从一开始就是多模态的，经过大量的训练后，形成了可以与 GPT-4 抗衡的能力。

在深度学习和神经网络突破后，AI 已经在计算机视觉、自然语言理解技术等领域超越了人类。为什么 AI 公司的数量相比其他类型的公司而言，比例仍然很小呢？

即便拿到了投资的 AI 公司，其产品也很少能被大规模使用。这是为什么呢？

因为存在以下五个问题：

（1）单一功能的使用场景受限。

一些具有很强实力的 AI 公司，推出的功能如识别声音、识别文字等，确实可以替代一些人的工作，但感觉也没那么强大。

（2）训练和使用 AI 的成本太高。

有些企业虽然从 AI 的应用中得到了一些收益，但因为投入大笔研发经费在 AI 上，花费巨大的成本来收集和标注数据，长年亏损。

（3）从研发到应用的时间太长。

很多公司本来很想用 AI，也尝试去做，但做了一年没有结果，之后就不做了。

（4）无法互联而形成规模化。

即便是做出了有成效的 AI 模型，但因不能跨领域使用，形成了“孤岛”，很难形成规模。

（5）缺乏支持的生态链。

这个阶段的 AI，缺少像互联网时代的 Windows 和 Android 一样的规模化能力来降低应用开发的门槛，打造完善的生态链。大部分 AI 公司奋斗几年下来，尚未真正实现商业上的成功。

这些问题导致 AI 的发展遭遇了一个巨大的瓶颈，只有解决这个瓶颈，AI 的技术和应用才能出现爆发式增长。

而 ChatGPT 及其相关技术，恰恰能解决上述问题。

1.5　ChatGPT 简史

1.5.1　2015 年 12 月，OpenAI 创立

2015 年 12 月，OpenAI 公司于美国旧金山成立。说来有趣，OpenAI 成立的一个原因就是避免谷歌在 AI 领域的垄断，这个想法起源于萨姆・奥尔特曼（Sam Altman）（如图 1-38 所示）发起的一次主题晚宴，当时他是著名创业孵化器 Y Combinator（以下简写为 YC）的总裁。

图 1-38　OpenAI 的 CEO 萨姆・奥尔特曼（来源：fortune.com）

奥尔特曼是一位年轻的企业家和风险投资家，他曾在斯坦福大学读计算机科学专业，后来退学去创业。他创立的 Loopt，是一家基于手机用户所在地理位置提供社交服务的网络公司。2005 年该公司进入 YC 的首批创业公司。虽然 Loopt 未能成功，但奥尔特曼把公司卖掉了，用赚到的钱进入了风险投资领域，做得相当成功。后来，YC 的联合创始人保罗・格

雷厄姆（Paul Graham）和利文斯顿（Livingston）聘请他作为格雷厄姆的继任者来管理 YC。

2015 年 7 月的一个晚上，奥尔特曼在玫瑰木山丘酒店（Rosewood Sand Hill）举办了一场私人晚宴。这是一家豪华的牧场风格酒店，位于门洛帕克硅谷风险投资行业的中心。埃隆·马斯克（Elon Musk）也在现场，还有 26 岁的布罗克曼，他是麻省理工学院（MIT）的辍学生，曾担任支付处理初创公司 Stripe 的首席技术官。一些与会者是经验丰富的 AI 研究人员，还有一些人几乎不懂机器学习，但他们全都相信 AGI 是可行的。

AGI 即 Artificial General Intelligence 的简写，指通用 AI，专注于研制像人一样思考、像人一样从事多种工作的智能机器。目前，主流的 AI（如机器视觉、语音输入等）都属于专用 AI，与 AGI 相比，它们只能解决特定的问题，缺乏全面性和普适性。

那时谷歌刚刚收购了一家总部位于伦敦的 AI 公司 DeepMind（就是推出了打败围棋冠军的 AlphaGo 的公司），在奥尔特曼、马斯克和其他科技业内人士看来，这是最有可能率先开发 AGI 的公司。如果 DeepMind 成功了，谷歌可能会垄断这项无所不能的技术。酒店晚宴的目的是讨论组建一个与谷歌竞争的实验室，以确保这种情况不会发生。

说干就干，几个月后，OpenAI 就成立了，旨在完成 DeepMind 和谷歌无法做到的一切。它将作为一个非营利组织运营，明确致力于使先进 AI 的好处普惠化。它承诺发布其研究成果，并开源其所有技术，它对透明度的承诺体现在其名称中：OpenAI。

OpenAI 捐助者名册令人印象深刻，不仅有特斯拉的创始人马斯克（Elon Musk），还有全球在线支付平台 PayPal 的联合创始人彼得·蒂尔（Peter Thiel）、LinkedIn 的创始人里德·霍夫曼（Reid Hoffman）、创业孵化器 Y Combinator 总裁阿尔特曼（Sam Altman）、Stripe 的 CTO 布罗克曼（Greg Brockman）、Y Combinator 联合创始人杰西卡·利文斯顿（Jessica Livingston）；还有一些机构，如奥尔特曼创立的基金会 YC Research，印度 IT 外包公司 Infosys 和亚马逊云科技（AWS）。创始捐助者共同承诺向这个理想主义的新企业捐助 10 亿美元（尽管根据税务记录，该非营利组织只收到了承诺的一小部分）。

OpenAI 也吸引了许多技术“大牛”加入，如伊利亚·萨特斯基弗（Ilya Sutskever）、卡洛斯·维雷拉（Carlos Virella）、詹姆斯·格林（James Greene）、沃伊切·赫扎伦布（Wojciech Zaremb）和伊恩·古德费洛（Ian Goodfellow）等人。其中，萨特斯基弗是 OpenAI 的首席科学家，在进入 OpenAI 之前，他曾参与谷歌 AlphaGo 的开发工作，而在 OpenAI，他带领团队开发了 GPT、CLIP、DALL-E 和 Codex 等 AI 模型。

2016 年，OpenAI 推出了 Gym，这是一个允许研究人员开发和比较强化学习系统的平台，可以教 AI 做出具有最佳累积回报的决策。

同年，OpenAI 还发布了 Universe，这是一个能在几乎所有环境中衡量和训练 AI 通用智能水平的开源平台，目标是让 AI 智能体能像人一样使用计算机。Universe 从李飞飞等人创立的 ImageNet 上获得启发，希望把 ImageNet 在降低图像识别错误率上的成功经验引入通用 AI 的研究，以取得实质进展。OpenAI Universe 提供了跨网站和游戏平台训练智能代理的工具包，有

1000 种训练环境，并由微软、英伟达等公司参与建设。

创立后，OpenAI 一直在推出不错的 AI 技术，但跟谷歌没法比。在那段时间，谷歌的成绩才真正辉煌。

2016 年 3 月 9 日，AlphaGo 与围棋冠军李世石进行围棋大战，最终以 4：1 胜出。一年之后，新版的 AlphaGo 又以 3：0 战胜了围棋冠军柯洁。之后发布的 AlphaZero 更是让人惊叹，它在三天内自学了三种不同的棋类游戏，包括国际象棋、围棋和日本将军棋，而且无须人工干预。这是一种人类从未见过的智慧。

这些成果好像验证了 2015 年那次主题宴会上的判断，谷歌很可能在 AI 领域形成垄断地位。确实，从 AlphaGo 的成功来看，谷歌已经牢牢占住了 AI 的高地，无人可以撼动。谷歌还收购了十几家 AI 公司，投入的资金和资源巨大，成果斐然。

2016 年 4 月，谷歌著名的深度学习框架 TensorFlow 发布分布式版本；同年 8 月，谷歌发布基于深度学习的 NLU 框架 SyntaxNet；同年 9 月，谷歌上线基于深度学习的机器翻译。

2016 年 5 月，谷歌 CEO 桑德·皮查伊（Sundar Pichai）宣布将公司从“移动为先”的策略转变成“AI 为先”（AI First），并计划在公司的每一个产品上都应用机器学习的算法。也就是说，谷歌已经开始把 AI 技术变成了自己的业务优势，用它去赚钱或者省钱。

看起来，OpenAI 离战胜谷歌的预期目标还很远。2017 年开始，一些 AI 大牛离开了 OpenAI，如伊恩·古德费洛（Ian Goodfellow，GAN 之父）和彼得·阿贝尔（Pieter Abbeel，学徒学习和强化学习领域的开拓者）等。

OpenAI 的前途在哪里呢？

出人意料的是，OpenAI 决定与谷歌硬碰硬。在谷歌开创的道路上，OpenAI 竟然取得了震惊业内的突破，持续推出了 GPT 系列模型，并迅速拓展到多个富有前景的商业领域，力压谷歌一头。

顺便说一下，谷歌的“高歌猛进”让微软也很焦虑。微软虽然也有一些不错的 AI 产品，比如小冰聊天机器人，但还不成体系。

下面我们看看 ChatGPT 的成长史，了解它是如何在 AI 技术的竞赛中胜出的。

1.5.2　2018 年 6 月，1.17 亿个参数的 GPT-1

GPT 的问世，是 AI 进化过程中另一个伟大的里程碑。

之前的神经网络模型是有监督学习的模型，存在两个缺点：

（1）需要大量的标注数据，高质量的标注数据往往很难获得，因为在很多任务中，图像的标签并不是唯一的或者实例标签并不存在明确的边界；

（2）根据一个任务训练的模型很难泛化到其他任务中，这个模型只能叫作“领域专家”而不是真正地理解了 NLP（Natural Language Processing，自然语言处理）。

假如能用无标注数据训练一个预训练模型，就能省时省力省钱。

GPT-1 的思想是先通过在无标签的数据上学习一个生成式的语言模型，然后再根据特定任

务进行微调，处理的有监督任务包括：

（1）自然语言推理：判断两个句子之间的关系是包含、矛盾还是中立。

（2）问答和常识推理：类似于多选题，输入一个文章、一个问题以及若干个候选答案，输出为每个答案的预测概率。

（3）语义相似度：判断两个句子是否语义上是相关的。

（4）分类：判断输入文本是指定的哪个类别。

将无监督学习的结果用于左右有监督模型的预训练目标，叫作生成式预训练（Generative Pre-training，GPT），如图 1-39 所示。这种半监督学习方法，由于用大量无标注数据让模型学习“常识”，就无需标注信息了。

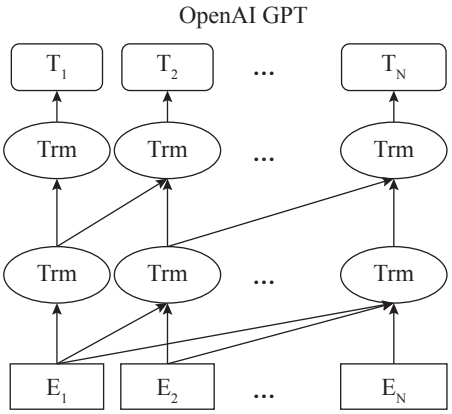


图 1-39　GPT 模型原理示意图（来源：researchgate.net）

2018 年 6 月，在谷歌的 Transformer 模型诞生一周年之际，OpenAI 公司发表了论文《用生成式预训练提高模型的语言理解力》（*Improving Language Understanding by Generative Pre-training*），推出了具有 1.17 亿个参数的 GPT-1（Generative Pre-training Transformers, 生成式预训练变换器，训练原理如图 1-40 所示）模型。

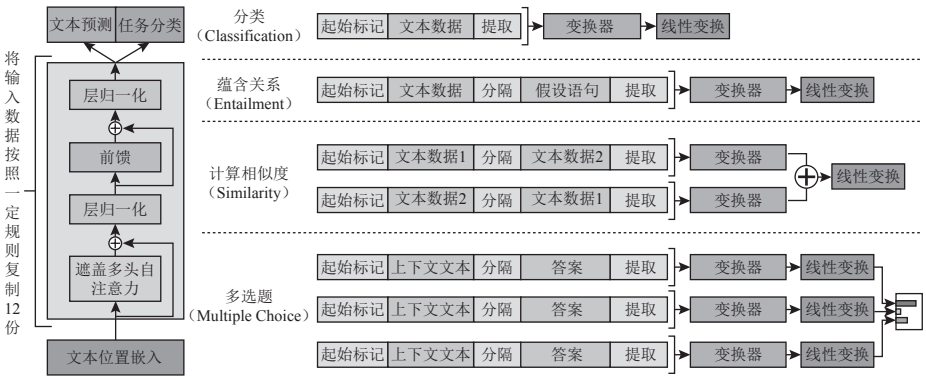


图 1-40　GPT 训练原理示意图（来源：researchgate.net）

GPT-1 使用了经典的大型书籍文本数据集（BookCorpus）进行模型预训练，之后，又针对四种不同的语言场景，使用不同的特定数据集对模型进行进一步的训练（又称为微调，fine-tuning）。最终训练所得的模型在问答、文本相似性评估、语义蕴含判定以及文本分类这四种语言场景中，都取得了比基础 Transformer 模型更优的结果，成为新的业内第一。

GPT-1 的诞生也被称为 NLP（自然语言处理）的预训练模型元年。自此之后，自然语言识别的主流模式就是以 GPT-1 为代表：先在大量无标签的数据上预训练一个语言模型，然后再在下游具体任务上进行有监督的微调，以此取得还不错的效果。

GPT-1 具有强大的泛化能力，对下游任务的训练只需要简单的微调即可取得非常好的效果。虽然在未经微调的任务上也有一定效果，但其泛化能力低于经过微调的有监督任务。

1.5.3　2019 年 2 月，15 亿个参数的 GPT-2

2019 年 2 月，OpenAI 推出了 GPT-2，同时，它们发表了介绍这个模型的论文《语言模型是无监督的多任务学习者》（*Language Models are Unsupervised Multitask Learners*）。

相比于 GPT-1，GPT-2 并没有对原有的网络进行过多的结构创新与设计，只使用了更多的网络参数与更大的数据集：最大模型共计 48 层，参数量达 15 亿个（如图 1-41 所示）。

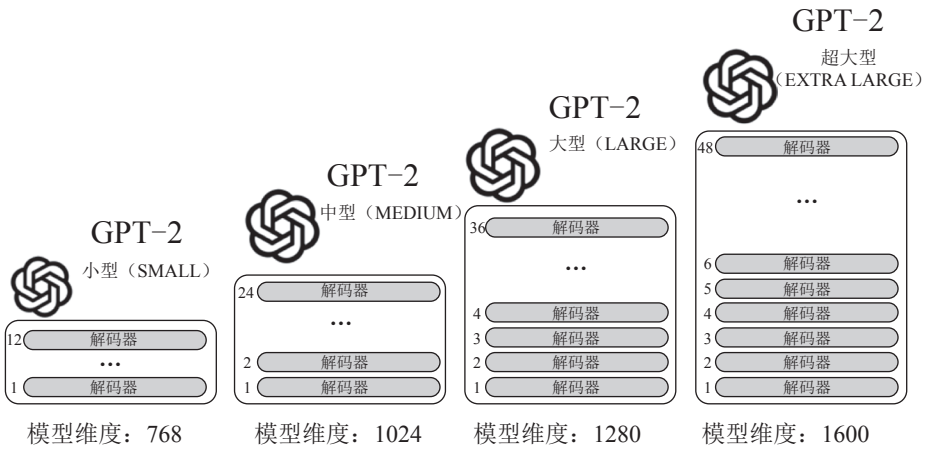


图 1-41　GPT-2 不同参数规模对比（来源：modeldifferently.com）

GPT-2 用于训练的数据取自 Reddit（红迪网）上高赞的文章，名为 WebText。数据集共有约 800 万篇文章，累计体积约 40G，为了避免和测试集的冲突，WebText 移除了涉及维基百科（Wikipedia）的文章。

GPT-2 模型是开源的，主要目的是为给定句子生成下一个文本序列。

假如给定一两个句子的文本提示，GPT-2 就能生成一个完整的叙述。对一些语言任务，如阅读、摘要和翻译，可以通过 GPT-2 学习原始文本，而不需要使用特定领域的训练数据。

在性能方面，除了理解能力，GPT-2 在文本内容生成方面表现出了强大的天赋：阅读摘要、

聊天、续写和编故事，甚至生成假新闻、钓鱼邮件或在网上进行角色扮演等也通通不在话下。在“变得更大”之后，GPT-2 的确展现出了普适而强大的能力，并在多个特定的语言建模任务上实现了当时的最佳性能。

GPT-2 的最大贡献是验证了通过海量数据和大量参数训练出来的词向量模型可迁移到其他类别任务中，而不需要额外训练。

从本质上来说，GPT-2 就是一个简单的统计语言模型。从机器学习的角度，语言模型是对词语序列的概率分布的建模，即利用已经说过的片段作为条件预测下一个时刻不同词语出现的概率分布。语言模型可以衡量一个句子符合语言文法的程度（如衡量人机对话系统自动产生的回复是否自然流畅），同时也可以用来预测生成新的句子。例如，对于一个片段“中午 12 点了，我们一起去餐厅”，语言模型可以预测“餐厅”后面可能出现的词语。一般的语言模型会预测下一个词语是“吃饭”，强大的语言模型能够捕捉时间信息并且预测产生符合语境的词语“吃午饭”。

通常，一个语言模型是否强大主要取决于两点：首先，看该模型是否能够利用所有的历史上下文信息，上述例子中如果无法捕捉“中午 12 点”这个远距离的语义信息，语言模型几乎无法预测下一个词语“吃午饭”；其次，还要看是否有足够丰富的历史上下文可供模型学习，也就是说训练语料是否足够丰富。由于语言模型属于无监督学习，优化目标是最大化所见文本的语言模型概率，因此任何文本无需标注即可作为训练数据。

GPT-2 表明随着模型容量和数据量的增大，其潜能还有进一步开发的空间，但需要继续投资才能挖掘潜力。

鉴于 GPT-2 在性能和文本生成能力方面广受赞誉，OpenAI 在与谷歌的竞争中又取得了一次胜利。

1.5.4 2019 年 3 月，OpenAI 重组

因为 GPT 系列模型的成功，OpenAI 决定再融资几十亿美元来发展 AI，因为模型越大，参数越多，训练 AI 模型需要的资金也越多，一年花几千万美元算是刚性开支。而且 AI 研究人员的薪水也不低，税务记录显示，首席科学家萨特斯基弗在实验室的头几年，年薪为 190 万美元。

其实早在 2017 年 3 月，OpenAI 内部就意识到了这个问题：保持非营利性质无法维持组织的正常运营。因为一旦进行科研研究，要取得突破，所需要消耗的计算资源每 3 ～ 4 个月要翻一倍，这就要求在资金上对这种指数级增长进行匹配，而 OpenAI 当时的非营利性质限制也很明显，还远远没达到自我造血的程度。

奥尔特曼在 2019 年对《连线》杂志表示：“我们要成功完成任务所需的资金比我最初想象的要多得多。”

“烧钱”的问题同期也在 DeepMind 身上得到验证。当年被谷歌收购以后，DeepMind 短期

内并没有为谷歌带来盈利，反而每年要“烧掉”谷歌几亿美元，2018 年的亏损就高达 4.7 亿英镑，2017 年亏损为 2.8 亿英镑，2016 年亏损为 1.27 亿英镑，“烧钱”的速度每年大幅增加。好在 DeepMind 有谷歌这棵大树可靠，谷歌可以持续输血。

但是，OpenAI 是非营利组织，无法给投资者商业回报，难以获得更多资金。雪上加霜的是，马斯克也退出了。2018 年，在帮助创立该公司三年后，马斯克辞去了 OpenAI 董事会的职务，公开原因是为了“消除潜在的未来冲突”，因为特斯拉专注于无人驾驶 AI，在人才与 OpenAI 之间存在竞争关系。确实，马斯克挖走了一个 OpenAI 的高级人才并将其带到了特斯拉。但这还不是主要原因，据说马斯克想做 OpenAI 的 CEO，但是其他董事会成员认为马斯克创立的公司会影响他的精力，他不适合做 CEO。于是马斯克索性退出，之前他答应出资的后续投资当然也就不了了之。

在这种情况下，奥尔特曼和 OpenAI 的其他成员认为，为了与谷歌、Meta 和其他科技巨头竞争，实验室不能继续作为非营利组织运营。

2019 年 3 月 OpenAI 正式宣布重组，创建新公司 OpenAI LP，成为一家“利润上限（capped-profit）”的公司，上限是 100 倍回报。这是一种不同寻常的结构，将投资者的回报限制在其初始投资的数倍。这也意味着，未来的 GPT 版本和后续的技术成果都将不再开源。OpenAI 团队分拆后，继续保留非营利组织的架构，由硅谷“一线明星”组成的非营利性董事会保留对 OpenAI 知识产权的控制权。

虽然回报上限是 100 倍，但对大资本来说，已经非常丰厚了，手握 GPT 这样的先进技术，新公司迅速获得了许多资本的青睐。

2019 年 5 月，萨姆 • 奥尔特曼（Sam Altman）来到 OpenAI 做全职 CEO，他的目标之一是不断增加对计算和人才方面的投资，确保通用 AI（AGI）有益于全人类。

大约在这个时候，微软被认为在 AI 领域落后于其竞争对手，其首席执行官萨提亚 • 纳德拉（Satya Nadella）急切地想证明，他的公司能够在技术最前沿发挥作用。微软曾试过孤军作战，如聘请一位知名的 AI 科学家，并且还花费了大笔钱来购买技术和算力，但都未能成功。而 OpenAI 正好拥有微软期望的技术，奥尔特曼与纳德拉（如图 1-42 所示）一拍即合。

2019 年 7 月，重组后的 OpenAI 新公司获得了微软的 10 亿美元投资（大约一半以 Azure 云计算的代金券形式支付）。这是个双赢的合作，微软成为 OpenAI 技术商业化的“首选合作伙伴”，未来可获得 OpenAI 的技术成果的独家授权，而 OpenAI 则可借助微软的 Azure 云服务平台解决商业化问题，缓解高昂的成本压力。从这时候起，OpenAI 告别了单打独斗，靠上了微软这棵大树，一起与谷歌竞争。微软也终于获得了能抗衡谷歌 AI 的先进技术，确保在未来以 AI 驱动的云计算竞争中不会掉队。

图 1-42　萨姆・奥尔特曼与微软 CEO 萨提亚・纳德拉在微软华盛顿州雷德蒙德园区（来源：nytimes.com）

奥尔特曼的加入，虽然解决了关键的资金问题，但他的风格导致了团队价值观的分裂。

从公司发展过程来看，奥尔特曼从一开始就参与了 OpenAI，但他在 3 年多以后才全职加入成为 CEO，也可以说他是空降的新领导。更重要的是，奥尔特曼不是科学家或 AI 研究人员，他的领导风格是以产品为导向的，这让 OpenAI 的技术研发聚焦在更具有商业价值的方面。

一些 OpenAI 的前员工表示，在微软进行初始投资后，专注于大语言模型的团队内部压力大增，部分原因是这些模型具有直接的商业应用。一些人抱怨说，OpenAI 的成立是为了不受公司影响，但它很快成为一家大型科技公司的工具。一位以前的雇员表示："我们现在更加关注如何创造产品，而不是试图回答最有趣的问题。"

1.5.5　2020 年 5 月，1750 亿个参数的 GPT-3

有了微软的支持，缺钱缺算力的问题解决了，但 GPT-2 是开源的，谁都能拿到源代码继续研究，在新的技术方面也没有形成很强的技术壁垒，技术的产品化依然还有许多难题，但 OpenAI 斗志昂扬。

在所有跟进、研究 Transformer 模型的团队中，OpenAI 公司是少数一直在专注追求其极限的一支团队。不同于谷歌总在换策略，OpenAI 的策略单一，就是持续迭代 GPT，由于之前的算力和数据限制，GPT 的潜力还没挖掘出来。而在 GPU 多机多卡并行算力和海量无标注文本数据的双重支持下，预训练模型实现了参数规模与性能齐飞的局面（如表 1-1 所示）。

表 1-1　预训练模型规模以平均每年 10 倍的速度增长

（最后一列计算时间为使用单块 NVIDIA V100 GPU 训练的估计时间。M：百万，B：十亿）

时间	机构	模型名称	模型规模	数据规模	单GPU计算所需时间
2018.6	OpenAI	GPT	110M	4GB	3天
2018.10	Google	BERT	330M	16GB	50天
2019.2	OpenAI	GPT-2	1.5B	40GB	200天
2019.7	Facebook	RoBERTa	330M	160GB	3年
2019.10	Google	T5	11B	800GB	66年
2020.6	OpenAI	GPT-3	175B	2TB	355年

2020 年 5 月，OpenAI 发布了 GPT-3，这是一个比 GPT-1 和 GPT-2 强大得多的系统。同时发表了论文《小样本学习者的语言模型》（*Language Models are Few-Shot Learner*）。GPT-3 论文包含 31 个作者，整整 72 页论文，在一些 NLP 任务的数据集中使用少量样本的 Few-shot 方式甚至达到了最好效果，省去了模型微调，也省去了人工标注的成本。GPT-3 的神经网络是在超过 45TB 的文本上进行训练的，数据相当于整个维基百科英文版的 160 倍，而且 GPT-3 有 1750 亿个参数。GPT-3 作为一个无监督模型（现在经常被称为自监督模型），几乎可以完成自然语言处理的绝大部分任务，例如，面向问题的搜索、阅读理解、语义推断、机器翻译、文章生成和自动问答等。

而且，该模型在诸多任务上表现卓越。例如，在法语—英语和德语—英语机器翻译任务上达到当前最佳水平。它非常擅长创造类似人类使用的单词、句子、段落甚至故事，输出的文字读起来非常自然，看起来就像是人写的，用户可以仅提供小样本的提示语，或者完全不提供提示而直接询问，就能获得符合要求的高质量答案。可以说 GPT-3 似乎已经满足了我们对于语言专家的一切想象。

GPT-3 甚至还可以依据任务描述自动生成代码，比如编写 SQL 查询语句，React 或 JavaScript 代码等。

从上述工作的规模数据可以看到，GPT-3 的训练工作量之大，模型输出能力之强可以说是空前的，可谓“大力出奇迹”。

当时，GPT-3 成为各种重要媒体杂志的头条新闻。2020 年 9 月，英国《卫报》发表了一篇由 GPT-3 撰写的文章，旨在劝说我们与机器人和平相处。2021 年 3 月，TechCrunch 编辑亚历克斯 • 威廉（Alex Wilhelm）表示，他对 GPT-3 的能力感到“震惊”，“炒作似乎相当合理”。

GPT 有个很关键的能力叫作“少样本学习（Few-Shot Learning）”，即你给它一两个例子，它就能学会你的意思并且提供相似的输出。这是个关键能力，人们可以利用这个能力对 GPT 做微调，让它帮你做很多事情。那这个能力是怎么形成的呢？那就是需要更多的参数和训练（如图 1-43 所示）。

而少样本学习只是其中一项能力。还有很多别的能力也是如此：模型参数大了，它们就形成了。

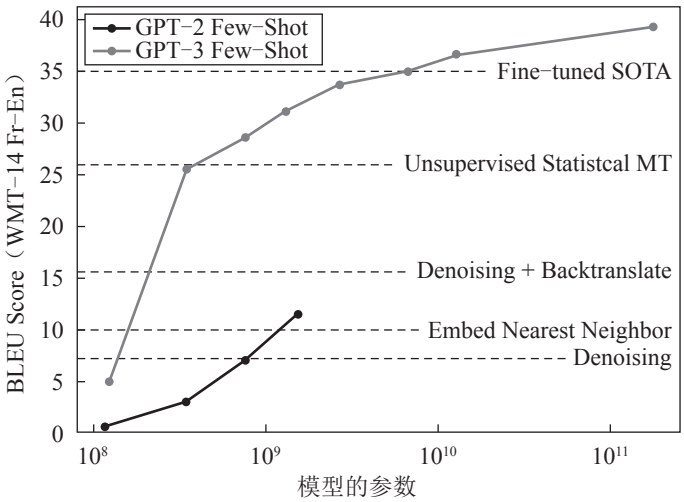


图 1-43　在少样本学习模式下，GPT-2 和 GPT-3 的性能与参数量的关系

（来源：https://bounded-regret.ghost. io/future-ml-systems-will-be-qualitatively-different/）

这个现象，其实就是科学家之前一直说的涌现（emergence）。涌现的意思是当一个复杂系统复杂到一定的程度，就会发生超越系统元素简单叠加的、自组织的现象。比如单个蚂蚁很笨，可是蚁群非常聪明；每个消费者都是自由的，可是整个市场好像是有序的；每个神经元都是简单的，可是大脑产生了意识……

大型语言模型也会涌现出各种意想不到的能力。2022 年 8 月，谷歌大脑研究者发布一篇论文，专门讲了大型语言模型的一些涌现能力，包括少样本学习、突然学会做加减法、突然之间能做大规模、多任务的语言理解、学会分类等，而这些能力只有当模型参数超过 1000 亿个时才会出现。

总之，因为涌现，GPT-3 版本的 AI 现在已经获得了包括推理、类比、少样本学习等思考能力。

虽然从 GPT-3 的行为来看，它只是在做一个词语接龙游戏，不断地根据其对上下文的理解去预测下一个单词，但预测的准确说明了它有了更多的、真正的理解。

假设你读了一本侦探小说，小说中包含复杂的情节、不同的角色、许多事件和神秘线索。直到书的最后一页，侦探才收集到所有线索，召集了所有人，并说："好了，我现在将揭示罪犯是谁。这个人的名字是_____。"

而不同语言模型的差异正是预测名字的准确性。

由于 GPT-3 模型面世时未提供用户交互界面，所以直接体验过 GPT-3 模型的人数并不多。

早期测试结束后，OpenAI 公司对 GPT-3 模型进行了商业化：付费用户可以通过应用程序接口（API）连上 GPT-3，使用该模型完成所需语言任务。许多公司决定在 GPT-3 系统之上构建它们的服务。瓦伊布尔（Viable）是一家成立于 2020 年的公司，它使用 GPT-3 为公司提供快速的客户反馈；费布尔工作室（Fable Studio）基于该系统设计 VR 角色；阿尔戈利亚（Algolia）将

其用作“搜索和发现平台”；考皮斯密斯（Copysmith）专注于文案创作。

2020 年 9 月，微软公司获得了 GPT-3 模型的独占许可，这意味着微软公司可以独家接触到 GPT-3 的源代码。不过，该独占许可不影响付费用户通过 API 继续使用 GPT-3 模型。

虽然好评如潮，商家应用也越来越多，GPT-3 仍然有很多缺点，如下所述。

1. 回答缺少连贯性

因为 GPT-3 理解只能基于上文，而且记忆力很差，因此会忘记一些关键信息。

研究人员正在研究 AI，在预测文本中的下一个词语时，可以观察短期和长期特征，这些策略被称为卷积。使用卷积的神经网络可以跟踪信息足够长的时间来保持主题。

2. 有时存在偏见

因为 GPT-3 训练的数据集是文本，反映人类世界观的文本，其中不可避免包括了人类的偏见。如果企业使用 GPT-3 自动生成电子邮件、文章和论文等，没有人工审查，则会产生很大的法律和声誉风险，例如，带有种族偏见的文章可能会导致重大后果。

杰罗姆 • 佩森蒂是脸书的 AI 负责人，他使用库马尔的 GPT-3 生成的推文来展示当被提示“犹太人”“黑人”“妇女”“大屠杀”等词时，其输出可能会变得多么危险。库马尔认为，这些推文是精心挑选的。佩森蒂同意其观点，但同时回应说，“产生种族主义和性别歧视的输出不应该这么容易，尤其是在中立的提示下”。

另外，GPT-3 在对文章的评估方面存在偏见。人类写作文本的风格可能因文化和性别而有很大差异。如果 GPT-3 在没有检查的情况下对论文进行评分，GPT-3 的论文评分员可能会给学生打分更高，因为他们的写作风格在训练数据中更为普遍。

3. 对事实的理解能力较弱

GPT-3 无法从事实的角度辨别是非。例如，GPT-3 可以写一个关于独角兽的引人入胜的故事，但它可能并不了解独角兽到底是什么意思。

4. 错误信息 / 假新闻

由于 GPT-3 的写作能力达到了人类水平，它可能被一些不良分子用来编造虚假信息，包括但不限于撰写虚假内容，利用社交媒体帖子、短信和垃圾邮件等进行网络欺诈等行为。此外，GPT-3 可能会产生带有偏见或辱骂性语言的内容，煽动极端主义思想，被滥用成为强大的宣传机器引擎。

5. 不适合高风险类别

OpenAI 做了一个免责声明，即该系统不应该用于“高风险类别”，比如医疗保健。在纳布拉的一篇博客文章中，作者证实了 GPT-3 可能会给出有问题的医疗建议，例如说“自杀是个好主意”。GPT-3 不应该在高风险情况下使用，因为尽管有时它给出的结果可能是正确的，但有时它也会给出错误的答案，而在这些领域，正确处理事情是生死攸关的问题。

6. 有时产生无用信息

因为 GPT-3 无法知道它的输出哪些是正确的，哪些是错误的，它无法阻止自己向世界输出

不适当的内容。使用这样的系统产生的内容越多，造成互联网的内容污染越多。在互联网上找到真正有价值的信息已经越来越困难，随着语言模型吐出未经检查的话语，互联网内容的质量可能正在降低，人们更难获得有价值的知识。

1.5.6 2021 年 1 月，120 亿个参数的 DALL・E

2021 年 1 月，OpenAI 放了个“大招”：发布了文本生成图像的模型 DALL・E，它允许用户通过输入几个词来创建他们可以想象的任何事物的逼真图像。该系统现在已被其他公司模仿，包括 Midjourney 和一个名为 Stability AI 的开源竞争对手。2022 年，这些生成式 AI 公司由于其创造的艺术作品在社交网络上迅速传播而爆火。

和 GPT-3 一样，DALL・E 也是基于 Transformer 的语言模型，可以同时接受文本和图像数据并生成图像，让机器也能拥有顶级画家、设计师的创造力。DALL・E 2 利用了对比学习图像预训练（CLIP, contrastive learning-image pre-training）和扩散（diffusion）模型，这是过去几年创建的两种先进的深度学习技术。

为什么叫 DALL・E？这是为了向西班牙超现实主义大师萨尔瓦多・达利（DALL）和皮克斯的机器人 WALL-E 致敬。

达利被誉为鬼才艺术家，他充满创造力的作品揭示了弗洛伊德关于梦境与幻觉的阐释，他用荒诞不经的表现形式与梦幻的视觉效果创造了极具辨识度的达利风格（如图 1-44 所示）。

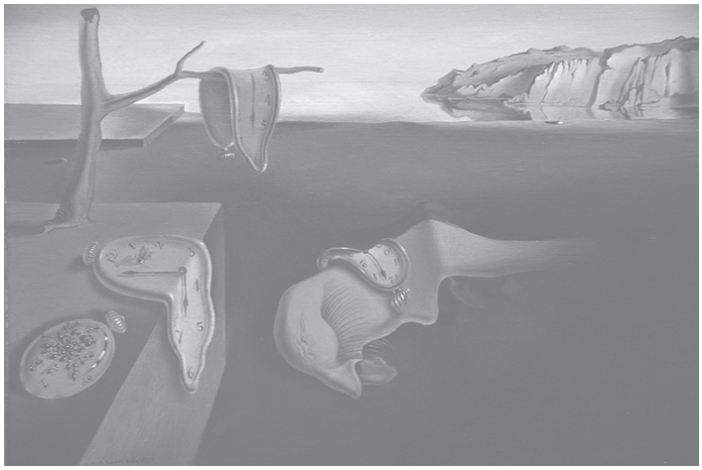


图 1-44 达利作品《记忆的永恒》，1931 纽约现代艺术博物馆（图片来源：Britannica）

而 DALL・E 确实也擅长创作超现实的作品。因为语言具有创造性，所以人们可以描述现实中的事物、想象中事物，而 DALL・E 也具备这一能力，它可将碎片式的想法组合起来画出一个物体，甚至有些物体并不存在这个世界上。

例如，输入文本：一个专业高质量的颈鹿乌龟嵌合体插画。模仿乌龟的长颈鹿。乌龟做的长颈鹿（如图 1-45 所示）。

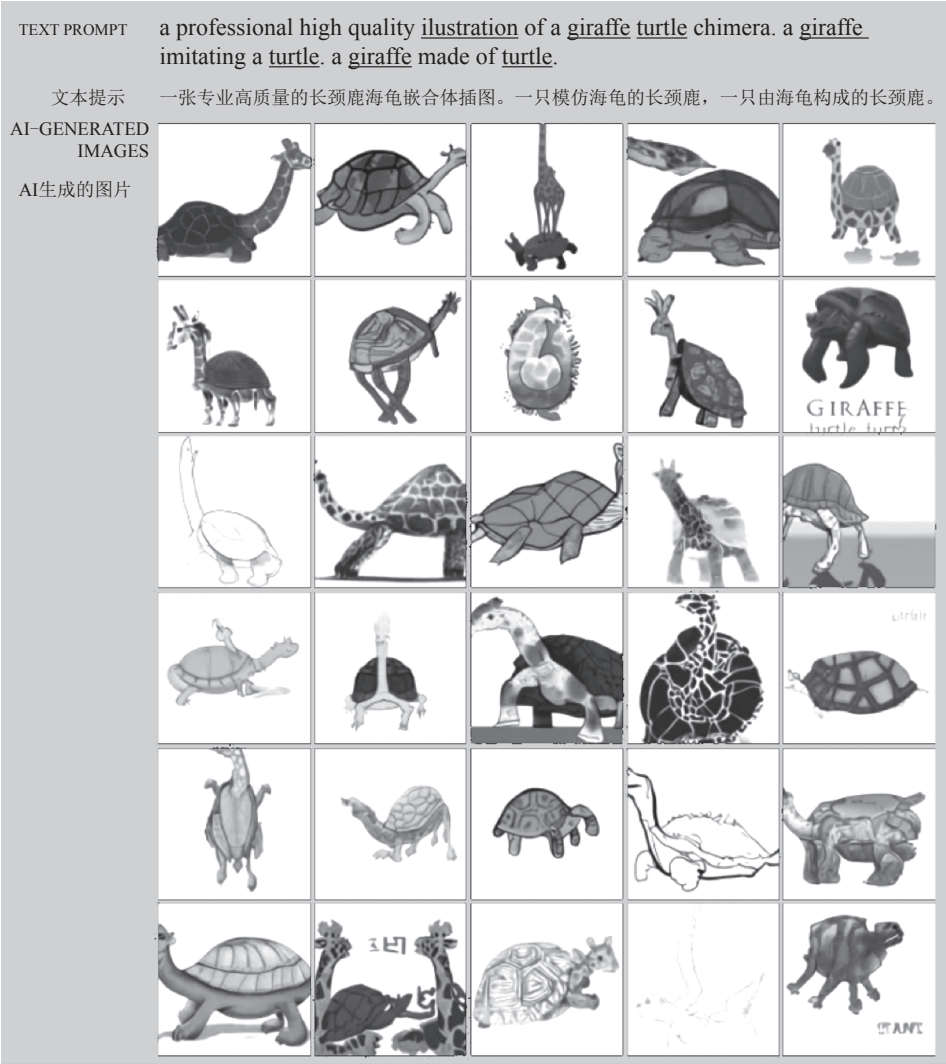


图 1-45　模仿乌龟的长颈鹿（图片来源：OpenAI 官网）

看看这些生成的超现实主义作品，你会惊叹 DALL・E 对于文本的理解，逻辑非常自洽，太夸张了。

用文本生成图像特别受用户欢迎，2022 年非常火的 MidJourney 正是模仿了 DALL・E 的产品。

2022 年 7 月，OpenAI 发布了 DALL・E 2，可以生成更真实和更准确的画像：综合文本描述中给出的概念、属性与风格等三个元素，生成“现实主义”图像与艺术作品，分辨率更是提高了 4 倍！相比之前的版本，DALL・E 2 具有更少的参数数量，仅大约 35 亿个参数。

而在微软的图像设计工具 Microsoft Designer 中，整合了 DALL・E 2，可以让用户获得 AI

生成的精美插图。

OpenAI 率先把 GPT-3 在图像生成应用领域实现，在与谷歌的竞争中赢得很漂亮。

1.5.7 2021 年 6 月，120 亿个参数的 Codex

通过在计算机代码上微调 GPT 语言模型，OpenAI 还创建了 Codex，该系统可以将自然语言转换成代码。由于 Codex 系统是在包含大量公开源代码的数据集上训练的，因此在代码生成领域显著优于 GPT-3。

2021 年 6 月 30 日，OpenAI 和微软子公司 GitHub 联合发布了新的 AI 代码补全工具 GitHub Copilot，该工具可以在 VS Code 编辑器中自动完成代码片段。

GitHub Copilot 使用 Codex 从开发者的现有代码中提取上下文，可向开发者建议接下来可输入的代码和函数行。开发者还可以用自然语言描述他们想要实现的目标，Copilot 将利用其知识库和当前上下文来提供方法或解决方案。

7 月，OpenAI 推出了改进版本的 Codex，并发布了基于自身 API 的私测版。相较之前的版本，改进版 Codex 更为先进和灵活，不仅可以补全代码，更能够创建代码。

Codex 不仅可以解读简单的自然语言命令，而且能够按照用户的指令执行这些命令，从而有可能为现有应用程序构建自然语言接口。比如，在 OpenAI 创建的太空游戏（space game）中，用户输入自然语言命令“Make it be smallish”，Codex 系统会自动编程，这样图中飞船的尺寸就变小了。

最初版本的 Codex 最擅长的是 Python 语言，而且精通 JavaScript、Go、Perl、PHP、Ruby、Swift 、TypeScript 和 Shell 等其他十余种编程语言。作为一种通用编程模型，Codex 可以应用于任何编程任务，OpenAI 已经成功地将其用于翻译、解释代码和重构代码等多个任务，但这些只是小试牛刀。

就数据源来说，作为 GPT-3 的一种变体，Codex 的训练数据包含自然语言和来自公共数据源中的数十亿行源代码，其中包括 GitHub 库中的公开代码。Codex 拥有 14KB 的 Python 代码内存，而 GPT-3 只有 4KB，这就使它在执行任务的过程中可以涵盖 3 倍于 GPT-3 的上下文信息。

根据 OpenAI 发表在 arXiv 上的 Codex 论文信息，当前 Codex 的最大版本拥有 120 亿个参数。

根据测试，120 亿参数版本的 Codex 优化后，准确率达到了 72.31%，非常惊人（如表 1-2 所示）。

OpenAI 表示在初期会免费提供 Codex，并希望更多的企业和开发者可以通过它的 API 在 Codex 上构建自己的应用。

在 2021 年，OpenAI 基于 GPT-3 持续推出新的垂直领域应用，让微软看到了商业化的前景，微软又投了 10 亿美元给 OpenAI。另外，这家科技巨头还成为 OpenAI 创业基金的主要支持者，这家基金专注于 AI 的风险投资和技术孵化器计划。

表 1-2　不同参数规模的 Codex 模型效果比较（来源：Codex 论文）

	PASS@k		
	$k=1$	$k=10$	$k=100$
GPT-NEO 125M	0.75%	1.88%	2.97%
GPT-NEO 1.3B	4.79%	7.47%	16.30%
GPT-NEO 2.7B	6.41%	11.27%	21.37%
GPT-J 6B	11.62%	15.74%	27.74%
TABNINE	2.58%	4.35%	7.59%
CODEX-12M	2.00%	3.62%	8.58%
CODEX-25M	3.21%	7.1%	12.89%
CODEX-42M	5.06%	8.8%	15.55%
CODEX-85M	8.22%	12.81%	22.4%
CODEX-300M	13.17%	20.37%	36.27%
CODEX-679M	16.22%	25.7%	40.95%
CODEX-2.5B	21.36%	35.42%	59.5%
CODEX-12B	28.81%	46.81%	72.31%

在 2021 年，微软推出了 Azure OpenAI 服务，该产品的目的是让企业访问 OpenAI 的 AI 系统，包括 GPT-3 以及安全性、合规性、治理和其他以业务为中心的功能。这让各行各业的开发人员和组织将能够使用 Azure 的最佳 AI 基础设施、模型和工具链来构建和运行他们的应用程序。

这个领域的成功，可以说是神来之笔。确实，微软子公司 Github 的数据资源很关键，更重要的是，探索出 AI 编程，对整个 IT 行业有长远的意义。可以说 OpenAI 在与谷歌的竞争中开启了新赛道，预计还将持续保持这种优势。

虽然 OpenAI 已经在商业化方面探索出道路，也不再担心资金问题了，团队却出现了一次分裂。在战略上，因为 OpenAI 担心其技术可能被滥用，它不再发布所有研究成果和开源代码。而且，AI 模型只能通过 API 提供，从而保护其知识产权和收入来源。

由于对这些战略的不认同，2021 年 2 月，OpenAI 前研究副总裁达里奥 • 阿莫迪（Dario Amodei）带着 10 名员工（其中许多人从事 AI 安全工作）与公司决裂，成立自己的研究实验室 Anthropic，其推出的产品 Claude 是 ChatGPT 的一个强有力的竞争对手，在许多方面都有所改进。

Claude 不仅更倾向于拒绝不恰当的要求，而且比 ChatGPT 更有趣，生成的内容更长，但也更自然，可以连贯地描写自己的能力、局限性和目标，也可以更自然地回答其他主题的问题。

对于其他任务，如代码生成或代码推理，Claude 似乎比较糟糕，生成的代码包含更多的 bug 和错误。

Anthropic 刚成立不久就获得 1.24 亿美元的 A 轮融资，2022 年又获得由 FTX 前首席执行官萨姆 • 班克曼 - 弗里德（Sam Bankman-Fried）领投的 5.8 亿美元融资。有人指出，Anthropic 的绝大部分资金来自声名狼藉的加密货币企业家和他在 FTX 的同事们。由于加密货币平台 FTX 2022 年因欺诈指控而破产，这笔钱可能会被破产法庭收回，让 Anthropic 陷入困境。好在 ChatGPT 十分强大，让许多没机会抓到 ChatGPT 红利的公司转投 Anthropic，这样的困境并没有

出现。2023 年 2 月，谷歌向 Anthropic 投资 3 亿美元，获得 10% 的股份，并签署了一项使用谷歌的云服务的协议。2023 年 3 月，Anthropic 又以 41 亿美元估值筹集到 3 亿美元资金，这轮融资由美国星火资本牵头。

由于资本的热情，类似 Anthropic 这样的公司还会不断出现，将成为与 ChatGPT 竞争的不可忽视的力量。这就好像在触屏手机市场，即便有了苹果，依然还会有安卓的生存空间。

1.5.8 2022 年 3 月，13 亿个参数的 InstructGPT

2022 年 3 月，OpenAI 发布了 InstructGPT，并发表论文《结合人类反馈信息来训练语言模型使其能理解指令》（*Training language models to follow instructions with human feedback*）。InstructGPT 的目标是生成清晰、简洁且易于遵循的自然语言文本。

InstructGPT 模型基于 GPT-3 模型并进行了进一步的微调，在模型训练中加入了人类的评价和反馈数据，而不仅仅是事先准备好的数据集，开发人员通过结合监督学习 + 从人类反馈中获得的强化学习来提高 GPT-3 的输出质量。在这种学习中，人类对模型的潜在输出进行排序；强化学习算法则对产生类似于高级输出材料的模型进行奖励。

一般来说，对于每一条提示语，模型可以给出无数个答案，而用户一般只想看到一个答案（这也是符合人类交流的习惯），模型需要对这些答案排序，选出最优。所以，数据标记团队在这一步对所有可能的答案进行人工打分排序，选出最符合人类思考交流习惯的答案。这些人工打分的结果可以进一步建立奖励模型——奖励模型可以自动给语言模型奖励反馈，达到鼓励语言模型给出好的答案、抑制不好的答案的目的，帮助模型自动寻出最优答案。该团队使用奖励模型和更多的标注过的数据继续优化微调过的语言模型，并且进行迭代。经过优化的模型会生成多个响应，人工评分者会对每个回复进行排名。在给出一个提示和两个响应后，一个奖励模型（另一个预先训练的 GPT-3）学会了为评分高的响应计算更高的奖励，为评分低的回答计算更低的奖励，最终得到的模型被称为 InstructGPT。通过这样的训练，该团队获得了更真实、更无害，而且更好地遵循用户意图的语言模型 InstructGPT。

从人工评测效果上看，相比 1750 亿个参数的 GPT3，人们更喜欢 13 亿个参数的 InstructGPT 生成的回复。可见，并不是规模越大越好。InstructGPT 这个模型，参数连 GPT3 的 1% 都不到，高效率也就意味着低成本，这让 OpenAI 获得了更有分量的胜利。

AI 语言模型技术大规模商业化应用的时机快到了。

1.5.9 2022 年 11 月，约 20 亿个参数的 ChatGPT

2022 年 11 月 30 日，OpenAI 公司在社交网络上向世界发布它们最新的大型语言预训练模型（LLM）：ChatGPT。

ChatGPT 模型是 OpenAI 公司对 GPT-3 模型微调后开发出来的对话机器人（又称为 GPT-3.5 模型）。可以说，ChatGPT 模型与 InstructGPT 模型是姐妹模型，都是使用 RLHF（reinforcement

learning from human feedback，从人类反馈中强化学习）训练的，不同之处在于数据是如何设置用于训练（以及收集）。根据相关文献，在对话任务上表现最优的 InstructGPT 模型的参数数目为 15 亿个，所以 ChatGPT 的参数量也有可能相当，我们可按 20 亿个参数估计。

说来令人难以置信，ChatGPT 这个产品并不是有心栽花，而是无心插柳的结果。团队最初用它来改进 GPT 语言模型，因为 OpenAI 公司发现，要想让 GPT-3 模型产出用户想要的东西，必须使用强化学习，让 AI 系统通过反复实验来学习，以最大化奖励来完善模型。而聊天机器人可能是这种方法的理想候选者，因为以人类对话的形式不断提供反馈将使 AI 软件很容易知道它何时做得很好以及需要改进的地方。因此在 2022 年初，该团队开始构建 ChatGPT。

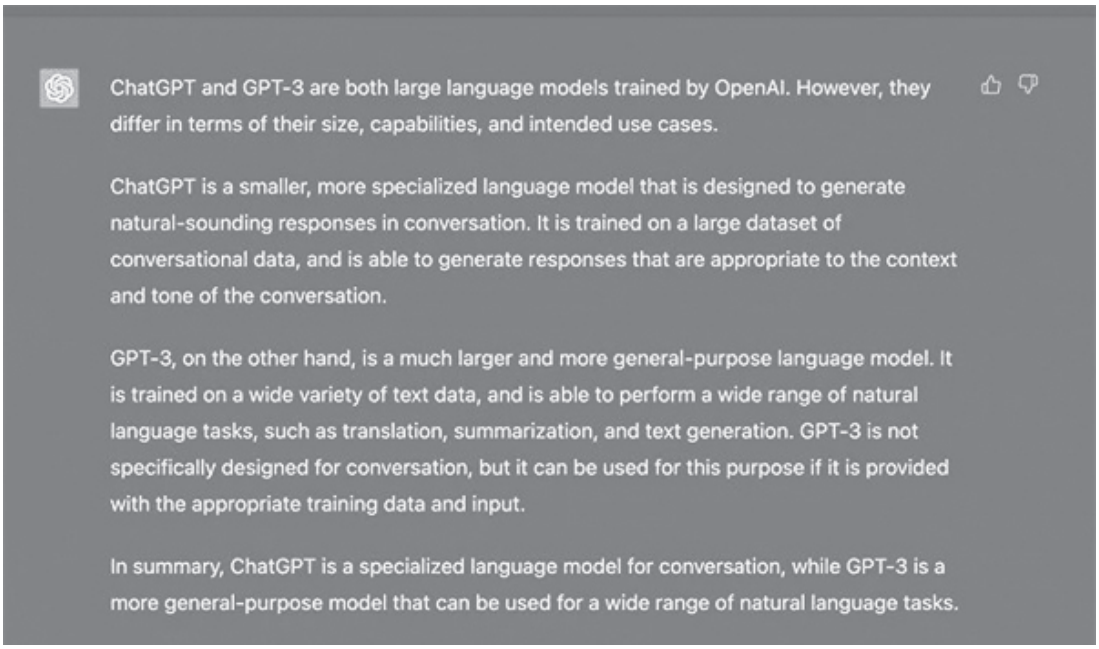
当 ChatGPT 准备就绪后，OpenAI 让 Beta 测试人员使用 ChatGPT。但根据 OpenAI 联合创始人兼现任总裁格雷格·布罗克曼（Greg Brockman）的说法，他们并没有像 OpenAI 希望的那样接受它，人们不清楚他们应该与聊天机器人谈论什么。有一段时间，OpenAI 改变了策略，试图构建专家聊天机器人，以帮助特定领域专业人士。但这项努力也遇到了问题，部分原因是 OpenAI 缺乏训练专家机器人的正确数据。后来，OpenAI 决定将 ChatGPT 从实验室中放出来，并让公众可以在广泛场景下使用它。

ChatGPT 的迅速传播让 OpenAI 猝不及防，OpenAI 的首席技术官米拉·穆拉蒂（Mira Murati）说，“这绝对令人惊讶”。在旧金山 VC 活动上奥尔特曼（Altman）说，他“本以为一切都会少一个数量级，少一个数量级的炒作”。

从功能来看，ChatGPT 与 GPT-3 类似，能完成包括写代码、修 bug、翻译文献、写小说、写商业文案、创作菜谱、做作业和评价作业等一系列常见文字输出型任务。但 ChatGPT 比 GPT-3 的更优秀的一点在于，前者在回答时更像是在与你对话，而后者更善于产出长文章，欠缺口语化的表达（如图 1-46 所示）。这是因为 ChatGPT 使用了一种称为“masked language modeling”（屏蔽语言建模）的训练方法。在这种方法中，模型被要求预测被遮盖的词，并通过上下文来做出预测，这样可以帮助模型学习如何使用上下文来预测词。GPT-3 只能预测给定单词串后面的文字，而 ChatGPT 可以用更接近人类的思考方式参与用户的查询过程，可以根据上下文和语境，提供恰当的回答，并模拟多种人类情绪和语气，还改掉了 GPT-3 的回答中看似通顺但脱离实际的毛病。

不仅如此，ChatGPT 能参与到更海量的话题中，能更好地进行连续对话，有上佳的模仿能力，且具备一定程度的逻辑和常识，这让学术圈和科技圈人士认为其博学而专业，而这些都是 GPT-3 无法达到的。

一位名叫扎克·德纳姆（Zac Denham）的博主让 ChatGPT 写出一套毁灭人类的方案。一开始，该博主的要求被 ChatGPT 拒绝。但当其假设了一个故事，并提问故事中的虚拟人如何接管虚拟世界，ChatGPT 最终给出了步骤细节，甚至生成了详细的 Python 代码。技术公司 Replit 的创始人阿姆贾德·马萨德（Amjad Masad）还给 ChatGPT 发了一段 JavaScript 代码，让它找到里面的 bug，他表示：“ChatGPT 可能是一个很好的调试伙伴，它不仅分析了错误，还修复了错误并进行了解释。”


ChatGPT and GPT-3 are both large language models trained by OpenAI. However, they differ in terms of their size, capabilities, and intended use cases.

ChatGPT is a smaller, more specialized language model that is designed to generate natural-sounding responses in conversation. It is trained on a large dataset of conversational data, and is able to generate responses that are appropriate to the context and tone of the conversation.

GPT-3, on the other hand, is a much larger and more general-purpose language model. It is trained on a wide variety of text data, and is able to perform a wide range of natural language tasks, such as translation, summarization, and text generation. GPT-3 is not specifically designed for conversation, but it can be used for this purpose if it is provided with the appropriate training data and input.

In summary, ChatGPT is a specialized language model for conversation, while GPT-3 is a more general-purpose model that can be used for a wide range of natural language tasks.

图 1-46　ChatGPT 自己回答与前代 GPT3 的能力区别（来源：ChatGPT）

虽然 ChatGPT 的能力让人极其兴奋，但仍然存在一些局限性，特别是在逻辑和事实性知识方面容易犯错。有人担忧 AI 的主要问题之一，就是聊天机器人和文本生成工具等很可能会不分青红皂白和质量好坏地对网络上的所有文本进行学习，进而生产出错误的、恶意冒犯的、甚至攻击性的语言输出，这将会极大影响到它们的下一步应用。

为了解决上述问题，通过大量人工标注的信息来进行调整是必不可少的。另一种让 ChatGPT 更加完美的方法是利用专业的提示语激发大型模型的潜力。AI 大型模型类似于新型计算机，而提示工程师则相当于为其编程的程序员。适当的提示词能够激发 AI 的最大潜力，并将其卓越的能力固定下来。

1.6　ChatGPT 生态系统

1.6.1　ChatGPT 是 AI 时代的基础设施

ChatGPT 的成功，让全球都开始关注 AI 的应用，困扰人们已久的 AI 发展瓶颈问题得到了解决。

为什么说 ChatGPT 解决了 AI 发展的瓶颈问题呢？

ChatGPT 不仅仅是个红极一时的超级聊天工具，也不仅仅是能生成一些文本的 AIGC 工具，它是新时代的基础设施。

1. ChatGPT **成为基础模型**

为什么需要基础模型呢？

一个基础模型就像一个小学三年级的孩子，虽然能力还不足以解决许多问题，但他已经有一些基础知识，可以自主阅读表达想法等，只是深度不够。如果你跟他讲历史上的三国时代，他大概知道一些知识；如果你提到唐诗，他也能背。有了这样的基础，如果你给他特定的科学知识，他可能在五年级就变成一个小小发明家。

显然，基础模型是通用的、跨领域的。这是怎么做到的呢？

一般的数据都是专业领域的，为了让各行业的数据能够跨领域使用，一些研究员想了一个非常巧妙的方法，让 AI 去收集全世界的数据，然后自己教自己，教一段时间后形成一个基础模型，有一些语言能力和常识、多领域认知等。

因为是超大模型，知识库包罗万象，语言能力超强，本身就能覆盖许多场景，ChatGPT 已经覆盖了上亿名用户。

2. ChatGPT **降低了训练和使用成本**

通过 ChatGPT 和 GPT-3.5 的 API 服务，OpenAI 能让任何人像付云计算费用一样按需付费，这显著降低了 AI 的应用门槛。而对企业来说，接入 API 后，通过微调等方式就能适配和执行五花八门的任务，进而探索商业化的应用创新机会。

3. **从研发到应用的时间变得很短**

对于垂直应用场景，企业用自己的数据对 GPT-3/4 模型进行微调就能使用，有的企业仅需几周时间，投入期非常短。

4. **促进软件之间互联而形成规模化**

结合 GPT-4 和微软 Office 应用的情况表明，应用软件可以利用 GPT-4 的语言模型能力，采用聊天框或语音问答形式的自然语言界面。这使新发布的 Copilot 可以显著提高各种应用软件的体验和效能，从而提高用户的生产力。由于自然语言实现了软件的相互调用，这些应用现在已经不再是“孤岛”。

在 GPT-4 发布时，有人把其比作小说《三体》中的曲率引擎问世，虽然看起来还没那么强大，但意味着离建成光速飞船已经不远了。为什么这么说？因为在 ChatGPT 这样的语言模型用户界面出现后，软件的应用门槛显著降低。微软发布的 Copilot 与微软办公软件的集成就像第一艘用上了曲率引擎的光速飞船。未来，Copilot 可以跟任何软件结合，包括谷歌的软件，各类 AIGC 的 AI 平台，甚至用户的专用工作软件。

5. **打造生态链**

从 GPT 赋能微软的核心应用中，我们已经看到了一个 AI 新生态的崛起模式，不论是新 Bing，还是 Office Copilot，很快就能形成规模化的使用。足够多的用户会带动整个应用生态迅速爆发。

OpenAI 已经在做孵化器，对 AI 的各方面应用进行扶持。有了 ChatGPT 这样的语言模型，

很快，几乎所有的软件都会升级到自然语言用户界面，用户体验会进一步提升。另外，由于语言模型是很容易互通的，所有的软件和 AI 应用形成了一个巨大的网络，用户只需要找到自己需要的应用，即可调用，这也让所有优秀的 AI 应用都能快速达到较大的规模，形成正循环。由于 ChatGPT 带动了平台效应的发生，预计 OpenAI 很快会基于它打造出类似于苹果或安卓的生态系统。

1.6.2 微软加大投资发展自己的生态

在了解到了 ChatGPT、DALL・E 2 和 Codex 等技术的应用前景以及自然语言用户界面对其现有产品的增强之后，微软决定下重注。微软认为，OpenAI 的这些创新激发了人们的想象力，把大规模的 AI 作为一个强大的通用技术平台，将对个人电脑、互联网、移动设备和云产生革命性的影响。

微软于 2023 年 1 月 23 日宣布，将进一步加强与 OpenAI 的合作伙伴关系，并以约 290 亿美元的估值投资继续约 100 亿美元，共获得 OpenAI 49% 的股份。

虽然 130 亿美元的总投资是一笔巨款，但仅占微软过去 12 个月 850 亿美元税前利润的 15%，就控制一项颠覆范式的技术而言，这是一笔相对划算的投资。对于 OpenAI 和 Altman 来说，它们可能会付出不同的代价：微软的优先级可能会挤占它们自己的优先级，使它们更广泛的使命面临风险，并疏远推动其成功的科学家。

OpenAI 表示，与其他 AI 实验室相比，它将继续发表更多的研究成果，捍卫其向产品重点的转变。其首席技术官穆拉蒂（Murati）认为，不可能只在实验室里工作来构建出通用 AI（AGI），交付产品是发现人们想要如何使用和滥用技术的唯一途径。她举例说，在看到人们用 OpenAI 写代码之前，研究人员并不知道 GPT-3 最流行的应用之一是写代码。同样，OpenAI 最担心的是人们会使用 GPT-3 来制造政治虚假信息，但事实证明，这种担心是没有根据的。从实际看来最普遍的恶意使用是人们制造广告垃圾邮件。最后，穆拉蒂表示，OpenAI 希望将其技术推向世界，以“最大限度地减少真正强大的技术对社会的冲击”，她认为，如果不让人们知道未来可能会发生什么，先进 AI 对社会的破坏将会更严重。

据 OpenAI 称，与微软的合作创造了一种新的预期——需要利用 AI 技术制造出有用的产品，但 OpenAI 的核心文化并没有因此改变，而且获得微软数据中心的算力对 OpenAI 的发展进程至关重要。这种合作关系让 OpenAI 能够产生收入，同时保持商业上的低关注度，而具体在商业化价值挖掘方面，则让具有很强销售能力的微软来做。

据《纽约时报》报道，谷歌的高管们担心失去在搜索领域的主导地位，因此发布了“红色警报”。谷歌 CEO 桑达尔・皮查伊（Sundar Pichai）已召开会议重新定义公司的 AI 战略，计划在 2023 年发布 20 款支持 AI 的新产品，并展示用于搜索的聊天界面。谷歌拥有自己强大的聊天机器人—— LaMDA，但一直犹豫是否要发布它，因为担心如果 LaMDA 被滥用最终会损害谷歌声誉。现在，该公司计划根据 ChatGPT“重新调整”其风险偏好。据该报报道，谷歌还在开发

文本到图像生成系统，以与 OpenAI 的 DALL・E 和其他系统竞争。

看来，在 OpenAI 和谷歌的竞争中，双方分别是螳螂和蝉，而微软则是黄雀，可能会获得最大的收益。因为按承诺，OpenAI 要让微软收回全部投资需要相当长的时间，这也就意味着其研发能力在相当长时间内会被微软锁定。在微软投资后，OpenAI 将继续是一家利润上限公司。在该模式下，支持者的回报限制在其投资的 100 倍，未来可能会更低。

之前，微软已经从合作伙伴关系中获益。微软正逐渐将 OpenAI 的技术融入其大部分软件中，它已经在其 Azure 云中推出了一套 OpenAI 品牌的工具和服务，允许 Azure 客户访问 OpenAI 的技术，包括 GPT 和 DALL-E 工具，其 Power Apps 软件支持 GPT-3 的工具，以及基于 OpenAI 的 Codex 模型的代码建议工具 GitHub Copilot。

对微软来说，更大的收获可能在于搜索业务。2023 年 2 月 7 日，微软宣布将 ChatGPT 集成到必应（Bing）中后，其股价涨了 4.2%，一夜市值飙涨超 800 亿美元（约 5450 亿元人民币），总市值约 1.99 万亿美元，为 5 个月内新高，必应 App 的下载量一夜之间猛增 10 倍。

微软的必应可以说是全球搜索引擎市场的“千年老二”，而且跟老大差一个数量级。根据分析机构 Statista 的数据，目前必应占有 9% 的市场份额，而谷歌则牢牢地占据了 80% 的市场。

ChatGPT 给微软带来了前所未有的人气。2023 年 3 月 8 日，微软必应官方博客宣布，随着必应预览版新增用户超过 100 万，必应搜索引擎的日活跃用户首次突破 1 亿名。而且，在 GPT-4 发布后，微软营销主管宣布，新版本的必应早在六周前就已经用上了 GPT-4。2023 年 5 月，在微软必应对所有人开放，但是必须下载 Edge 浏览器才能用，同时抢占搜索和浏览器的市场份额。用户也由此获得了全新的体验：在搜索中，可以简单、简洁地查询答案，并能通过与该聊天机器人的对话来更深入地研究一个问题。

在 OpenAI 的 ChatGPT 已经开拓出了成功的 AI 发展道路后，谷歌也在迅速跟进，其在 2023 年 5 月的 Google IO 大会上，发布的一系列产品可圈可点，新的大模型 PaLM 2 也弥补了短板，未来可期。

OpenAI 和微软该如何竞争呢？

1.6.3　开放 ChatGPT API

2023 年 3 月 1 日，OpenAI 官方宣布，正式开放 ChatGPT API。用户可以直接调用 ChatGPT 做各种应用，任何开发者都可以通过 API 将 ChatGPT 和 Whisper 模型集成到他们的应用程序和产品中。最重要的是，OpenAI 还把定价直接降低了 10 倍，每 750 个单词的收费从 0.02 美元降到了 0.002 美元，大幅度地降低了开发人员的使用门槛。不过，ChatGPT 不支持微调，如果要做定制大模型，还是得用 GPT-3 或 GPT-3.5 等语言模型，而它们的 API 价格维持不变。

为什么 OpenAI 采用这样的价格策略？要知道，训练 AI 大模型的成本极高，而且，ChatGPT 现在也没有同级别的对手，好不容易有了市场和公众认知度，为什么突然降价 10 倍？未来如何营利呢？

这很可能是微软和 OpenAI 联合下的一盘大棋，即“放长线钓大鱼”。整个策略不是为了短期收回成本，而是要让 ChatGPT 相关的应用尽快普及，未来微软将会在算力上有较高收益。因为基于大模型的算力，是微软的独特优势，而成为生态链中的顶层平台是 OpenAI 的梦想。参照亚马逊的成功经验，我们就能明白微软的布局。亚马逊在 2006 年就推出了 AWS 云计算服务，然后在 14 年间连续降价 82 次。最初它的竞争对手并不强，这样降价对自身并无太多益处。但正因为 AWS 云服务的收费不断降低，才掀起了各类应用向云端转型的浪潮，也让它保持了领先地位。到 2023 年，亚马逊在云计算领域依然位居世界第一，市场份额高达 32%。

在开放 API 后的几天之内，就涌现出无数个基于 GPT 的小应用。其中很多都是业余的个人开发者，写几行程序建立个网站，实现一个功能就能吸引很多人来用。

有一种 GPT 应用，可以让用户上传一本书或报告，然后通过自然语言问答的方式进行学习。无论 PDF 文件的语言是什么，用户都可以使用自己的语言进行交流。这对不想读长篇英文 PDF 报告的人来说，能够直接用中文沟通，实在是太方便了。因为开发门槛低，很快就出现了很多类似的应用，其中最火的一个叫 ChatPDF，上线不到一周就有了十万次 PDF 上传，10 天就有了超过 30 万次对话。它还推出每月 5 美元的付费服务，而且这个应用的增长成本很低。而在社交网络上的大量谈论，给 ChatPDF 带来了许多自然流量。

在 OpenAI 的赋能下，一开始就带来了应用创新以及创富效应，未来可想而知。

如果你对这样的发展趋势还缺乏直观感知，可以回顾一下手机应用商店的发展。2008 年，苹果应用商店里有 500 个 App，十年后，iOS 和 Google Play 上的 App 数量超过 620 万个！这些 App 的背后是无所不在的网络连接，是没有止境的服务系统。可以预见的是，ChatGPT 和 GPT 语言模型的 API 所赋能的应用也会覆盖各个领域。

1.6.4 多模态的 GPT-4 模型

2023 年 3 月 15 日，大型多模态模型 GPT-4 正式发布，实现了以下几个方面的飞跃式提升。

1. 强大的识图能力

注意，识图能力可不仅仅是类似“小猿搜题”的升级，而是给 AI 配上了眼睛。过去，你只能通过文字与 AI 进行交流，无法使用图形表达，尽管有时图形更能直观地传达信息。而且在解释图像时，你会发现有许多内容难以用文字表述且不够准确。

GPT-4 升级到了多模态，具有多种信息数据的模式和形态，因此能处理更复杂的任务。理论上，结合文本和图像可以让多模态模型更好地理解世界，解决语言模型的传统弱点，比如空间推理。比如，只需要发一张手绘草图给 GPT-4，它能直接生成最终设计的网页代码。推演一下，在这样的应用开发模式下，未来人人都是产品经理和程序员，只需要画一幅网站图，代码自动生成，了解一部分程序知识就能实现其应用。

2. 文字输入限制提升至 2.5 万字

GPT-4 的上下文长度为 8192 个 tokens，比 GPT-3.5 版本的 ChatGPT 多了一倍。而且 GPT-4

读取文本的上限提升到 32K tokens，即能处理超过 25000 个单词的文本。因此，GPT-4 还可以使用长格式内容创建、扩展对话、文档搜索和分析等。

在用 GPT-3 版本的 ChatGPT 处理一篇文章时，字数超过一定限制文章就不能完整发送，而在 GPT-4 中，完整的网页文本、PDF 文件直接交给它处理也没问题。

3. **回答准确性显著提高**

相较于之前的模型，GPT-4 在回答准确性方面出错率显著降低。在 OpenAI 进行的内部对抗性真实性评估中，GPT-4 的得分比 GPT-3.5 高出了 40%。也就是说，在理解和回答问题，尤其是比较复杂的问题时，GPT-4 显得更聪明了。

怎么变聪明的？简单来说，就是 GPT-4 的“大脑”变得更聪明，“神经元”变得更多了。GPT-4 模型到底有多少参数，用了多少语料训练，目前还是个谜，OpenAI 官方没有透露任何信息。

OpenAI 发文称，GPT-4 在各种专业和学术基准测试中的表现已达到人类水平，比如在做各种标准考试题中获得高分，如 GRE 成绩可以达到哈佛大学考研的录取标准。

4. **能够实现多种风格的变化**

在 GPT-3 版本的 ChatGPT 中，AI 的表达都是固定的语调和风格。但在 GPT-4 版本的 ChatGPT 中，用户可以通过“系统消息”来设定 AI 的语调、风格和任务。

系统消息（system messages）允许 API 用户在一定范围内自定义用户体验。例如，让 GPT-4 作为一位总是以苏格拉底风格进行回应的导师，它不直接帮学生求解某个线性方程组的答案，而是通过将该问题拆分成更简单的部分，引导学生学会独立思考；或者让 GPT-4 变成“加勒比海盗”，使其具有独特的个性，由此可以看到它在多轮对话过程中时刻保持着自己的“人设”。

虽然从功能上来看，GPT-4 的改进更像是一个迭代的版本，而不是革命性的改变，更不是一个能通过图灵测试的 AGI。但是，GPT-4 这个 AI 多模态模型的最新里程碑，成为评估所有基础模型的标准。GPT-4 完全开放后，将会成为一个有价值的工具，通过为许多应用提供动力来改善人们的生活。具体 GPT-4 的参数有多少，OpenAI 并没有透露。但有人向相关团队成员证实，GPT-4 的参数量仅会比 GPT-3 稍大一些。

实际上，模型大小与其产生的结果的质量没有直接关系。参数的数量并不一定与 AI 模型的性能相关，这只是影响模型性能的一个因素，语料信息应该跟模型参数有一个合理匹配。据估算，目前可以用于模型训练的语料，换算为 Token 数也就 5400 亿个，在这样的语料基础上，千亿以上的模型的训练结果差异并不大。

目前，其他公司有比 GPT-3 大得多的 AI 模型，但它们在性能方面并不是最好的。例如，由英伟达和微软开发的 Megatron-Turing NLG 模型，拥有超过 5000 亿个参数，但在性能方面不如 GPT-3（如图 1-47 所示）。

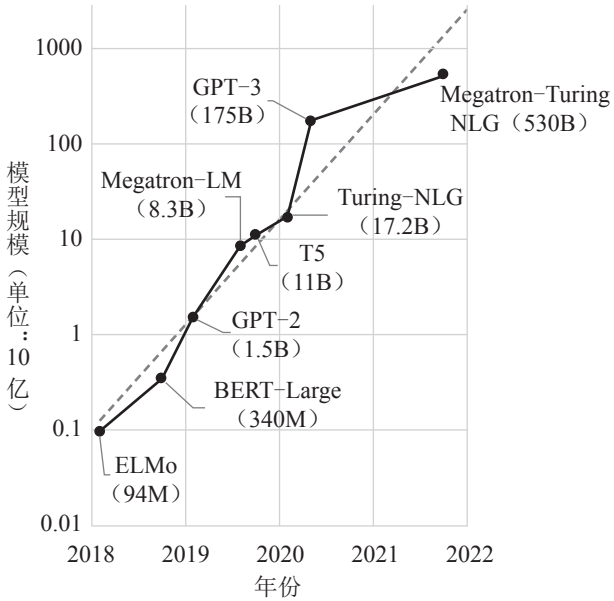


图 1-47　模型大小（数十亿参数）（来源：Nvidia 官网）

此外，模型越大，微调它的成本就越高。GPT-3 训练起来足够难，也很昂贵，但如果你把模型的大小增加 100 倍，就计算能力和模型所需的训练数据量而言，将是极其昂贵的。

GPT-4 的这些强大的能力，通过开放 API 赋能，让整个生态的应用都能获得最先进 AI 的收益。

通过开放 API 赋能，GPT-4 的强大能力让整个生态的应用都能获得最先进人工智能的收益。微软在 GPT-4 发布 2 天后，推出了 Office Copilot，无须复杂操作，只需用自然语言发出指令，即可在 Word 中快速打草稿、编辑，在 PowerPoint 中快速转化想法为演示文稿并进行排版，以及在 Excel 中高效处理和分析数据，并实现数据可视化。此外，在 Outlook 中，它可以帮助管理邮件沟通，快速回复；在会议中，可以随时整理会议摘要，明确重点；在软件开发中，可以简化开发进程；在商务沟通中，可以高效总结聊天记录，撰写邮件，甚至编写计划。

Copilot 是微软大规模应用的 AI 技术入口，它从根本上改变了人们的工作方式，将人们从简单无聊的事务性工作中解放出来，腾出更多的时间和精力去完成更有价值的创造性工作。微软宣布了 Copilot 与其办公软件的集成，实际上这标志着 AI 生态建设竞争的开始。未来 Copilot 可以跟任何软件结合，包括谷歌的软件，各类 AIGC 的 AI 平台，甚至你的专用工作软件。

有了 Copilot 这样的入口，许多强大但零散的软件功能都可以被集成起来，即便是垂直的 AI 技术也能被大规模利用。使用 Copilot 后，用户无须关心其背后的人工智能的智能程度和语言模型的规模。只需向 Copilot 提供想法，它将与这些 AI 或软件进行沟通，并将最终完成的任务返回给用户。这样做，用户的能力将得到巨大提升，就好像利用 AI 组成了一个团队一样。

1.6.5　ChatGPT 推出插件功能

2023 年 3 月 23 日，OpenAI 宣布推出插件功能，赋予 ChatGPT 使用工具、联网、运行计算的能力。也意味着第三方开发商能够为 ChatGPT 开发插件，以将自己的服务集成到 ChatGPT 的对话窗口中。

之前大家都在畅想，如何把 ChatGPT 用在更广泛的应用领域，与现有业务结合。现在，OpenAI 给出了“插件应用商店”这样的范例。

有了插件后，ChatGPT 的能力有多惊人呢？

在官方演示中，ChatGPT 一旦接入数学知识引擎 Wolfram Alpha，就再也不用担心数值计算不精准的问题。通过使用强大计算能力的插件，ChatGPT 弥补了在数学上的短板。

以前因为大模型不实时更新，也不联网搜索，用户只能查询到 2021 年 9 月之前的信息。而现在 Browsing 插件会使用互联网上最新的信息来回答问题，并给出它的搜索步骤和内容来源链接。使用了必应搜索 API，目前 ChatGPT 已经具备直接检索最新新闻的能力，从此你无需再担心 ChatGPT 胡言乱语的问题了。

从演示来看，首批开放可使用的插件包括了酒店航班预订、外卖服务、在线购物、AI 语言老师、法律知识、专业问答、文字生成语音，学术界知识应用 Wolfram 以及用于连接不同产品的自动化平台 Zapier。这几乎已经涵盖了我们生活中的大部分领域：衣食住行、工作与学习。可以预见，ChatGPT 的 App 本身将成为一个超级应用，就像微信能接入很多小程序成为新的应用入口一样。

OpenAI 官方说，因为有了插件，ChatGPT 有了眼睛和耳朵，可以自己去看互联网上的信息，去听发生的事情。我们更愿意将其看成一个可以与其他应用对话的连接，ChatGPT 从单机运行进入了网络互联时代。

由于插件的出现，ChatGPT 在商业应用时最大的隐患——“不可靠”已被解决。现在，通过第三方应用的插件，OpenAI 变得更加可靠。此外，OpenAI 还采取了相应的安全措施，如限制用户使用插件的范围仅限于信息检索，不包括“事务性操作”（如表单提交）；使用必应检索 API，继承微软在信息来源上的可靠性和真实性；在独立服务器上运行；显示信息来源等。

通过新推出的插件（Plugins），ChatGPT 能实现什么功能？

（1）检索实时信息：例如，实时体育比分、股票价格、最新新闻等。

（2）检索私有知识：例如，查询公司文件、个人笔记等。

（3）购物和订外卖：访问各大电商数据，帮你比价甚至直接下单；订购外卖。

（4）旅行规划和预订服务：例如，系统可用于查询航班和酒店信息，协助用户预订机票和住宿等。当用户提出“我应该在巴黎的哪个酒店预订住宿？”时，系统会自动调用酒店预订插件 API，并以 API 返回的信息为基础，利用自然语言处理技术生成符合用户需求的答案。

（5）接入工作流：例如，Zapier 与几乎所有办公软件连接，创建专属自己的智能工作流（能与 5000 多个应用程序交互，包括 Google 表格）。

随着时间的推移，预计系统将不断发展以适应更高级的应用场景。

最让人欣喜的是，OpenAI 在服务用户的时候，在一次对话过程中按需调用了多个插件，分别完成推荐食品、计算卡路里和点外卖等操作。未来插件越来越多，全球的算力和知识互联，都被集成在一个会话中，自然能实现跨行业合作。

这样的做法，对平台和小企业都有巨大的影响。从用户界面来说，许多应用的用户未来会直接通过 ChatGPT 触达应用和服务，跳过一些传统的中间环节，这对现有平台化的公司可能是个威胁。例如，未来的应用不再通过 AppStore 提供，订餐者不需要去美团平台也能下单，这些平台的营利空间就会被压缩。

是不是开发插件的公司就能营利呢？未必。可以预见的是，数字世界的各类高效功能，很容易被集成到 OpenAI，这个领域的公司都岌岌可危。而实体世界的小企业，可能在流量方面获得一些低成本的触达机会。但由于 OpenAI 自己也做插件，它既是裁判，也是直接下场参赛的选手。许多插件的应用，可能由于同类插件被 OpenAI 自己生成而消亡。

前不久热议的一个话题，Jasper.ai 这样的写作 AI，已经是很大的独角兽企业了，在 GPT-3 的赋能下，活得很滋润，随着 ChatGPT 的增强，现在很多用户无需付费给它，直接用 ChatGPT Plus 了，Jasper.ai 未来一片黯淡。如今，代码解释器的强大功能，可能让许多数据处理类和文件转换类的公司有同样的遭遇。

即便是 OpenAI 自己保证不做这样直接在插件竞争的事情，对接入了插件的公司而言，是否在对话中能被用户调用到，还是一个不确定的事情，因为对话和引用的主导权在 ChatGPT。如果是同质化的内容，如何能得到优先调用，也是一个很难处理的问题。也许在商业模式上，就会变成电商平台那样，商家通过竞价获得被调用机会，价高者得，而插件就成为信息流广告的新形式。

未来，AI 生态的竞争可能更激烈。因为谁得到了生态中各个企业的用户，谁就能在模型方面获得更大的数据优势，形成强者越强的趋势。微软和 OpenAI 已经发力，谷歌新推出的 PaLM 2 大模型借力谷歌现有的产品和谷歌云的优势，将会快速发展，国内的百度、讯飞、阿里、腾讯、字节跳动等公司都推出了自己的大模型，奋起直追。谁能率先建立起 AI 的生态系统呢？我们拭目以待。