
# 吴恩达传

## 人工智能教育与大众化的关键人物

金京镇律师

AI Library - kimkj.com - 2026

# 目录

# 从伦敦到斯坦福

1976年4月18日,英国伦敦,一个孩子出生了。这个孩子被取名为吴恩达(Andrew Yan-Tak Ng),他的父母是来自香港的移民。正是这样的家庭背景,让他同时拥有香港裔英国人、以及日后香港裔美国人的身份。父母具体从事什么职业、出于什么原因决定前往英国,现存的资料都无法确认。关于他幼年性格或内心冲突的记录也没有流传下来。本章不会用想象填补这些空白,只依循已确认的事实,如实追溯一个人从伦敦出生、经由亚洲、穿过美国几所大学、最终抵达斯坦福的轨迹。

幼年的吴恩达,成长期的大部分时间都在香港和新加坡度过。可以推测,往返两地成长的经历培养了他的国际视野,但没有具体的轶事可以佐证这一点。他何时第一次接触电脑、出于什么契机被数学和科学吸引,文献中也没有记载。不过,有一个明确的里程碑。1992年,他毕业于新加坡名校莱佛士书院。莱佛士书院是新加坡历史悠久、声誉卓著的学校,从这里毕业的学生中,有不少人后来在各自领域声名鹊起。吴恩达也加入了这一谱系,但在校期间哪门课影响了他的人生方向、遇到过哪位老师或朋友,都没有留下记录。因此,这段时期只以寥寥几句事实的形式留存下来:1976年生于伦敦,在香港与新加坡成长,1992年毕业于莱佛士书院。这三句话,就是他二十岁之前的人生中能够确认的全部内容。

离开新加坡后,吴恩达前往美国宾夕法尼亚州的卡内基梅隆大学。卡内基梅隆在计算机科学领域是世界顶尖的学府,后来更被人工智能研究者称为“摇篮”,是这一领域人才辈出的地方。吴恩达在这里同时修读计算机科学、统计学与经济学三个专业,成为一名三专业学生。这三门学科各自要求不同的思维方式:计算机科学训练人把问题拆解为算法,统计学培养从不确定的数据中读出规律的眼光,经济学则要求思考在有限资源下如何做出选择。一个人同时消化这三个专业绝非轻松的课业安排,但吴恩达完成了全部课程,并于1997年以全班第一名的成绩获得学士学位。他本人为何偏偏选择了这三个专业、本科阶段哪个瞬间决定了他的学术方向,文献中都没有留下他自己的回顾。不过从结果来看,这一组合为他日后成长为一名兼具数学严谨性、统计直觉与实用判断力的机器学习研究者,打下了必要的基础。

与本科学业重叠的1996年至1998年间,吴恩达获得机会在AT&T贝尔实验室担任研究员。贝尔实验室是二十世纪后半叶通信与计算机科学领域诞生诸多革新的地方,晶体管的发明,以及奠定信息论基础的研究,都出自这里。在这样的实验室里,吴恩达专注的课题是强化学习、模型选择与特征选择。强化学习是机器学习众多方法论中的一种,它不是由人一一告知正确答案,而是让系统通过反复试错自行找到更好的行动方式。这与训练狗时“做出想要的行为就给奖励,否则不给”的原理相似,只是被数学化地精细打磨过。模型选择是从能够解释给定数据的多个候选模型中,判断哪一个更可靠;特征选择则是从众多变量中挑出真正重要的少数几个。一个刚过二十岁的年轻人处理这三个课题,这件事本身就预示了他日后职业生涯的走向。这段经历是否是把他引向正式人工智能研究道路的决定性契机,本人并没有留下直接的回顾。但这一时期处理的课题,后来又作为贯穿他博士学位论文的核心概念重新出现,由此可见,贝尔实验室的这几年,正是他学术根基所在之处。

从卡内基梅隆毕业后,吴恩达转往马萨诸塞州剑桥的麻省理工学院,即MIT。他在这里于1998年获得理学硕士学位。在MIT就读期间,他留下的值得关注的成果并非一篇论文,而是一套软件系统。他亲手搭建了一个能够自动查找并索引网络上散布的机器学习研究论文的搜索引擎。如今论文检索服务已不稀奇,但在互联网草创的当时,能够自动收集并整理散落学术资料的系统本身十分罕见。吴恩达做出的这个搜索引擎,是最早一批可供公开使用、具备自动索引功能的系统之一,并成为后来在学界广泛使用的大型学术搜索

引擎CiteSeer(后更名为ResearchIndex)的先驱。他为何下决心制作这个系统、开发过程中遇到了怎样的技术难关,都没有留下记录。不过这段轶事,让人得以窥见吴恩达作为研究者的一种倾向:他不满足于让理论停留在理论层面,而是关注如何把它变成真正能用的工具。这种倾向在此后数十年间反复出现。

在MIT完成硕士学业后,吴恩达再度西行,前往加州大学伯克利分校攻读计算机科学博士学位。他的导师是机器学习领域享誉世界的迈克尔·艾·乔丹(Michael I. Jordan)。乔丹被公认为在统计学与机器学习之间架起桥梁的学者,他门下培养出的学生中,有多人后来成长为引领这一领域的研究者。吴恩达的博士学位论文答辩委员会中,除乔丹外,还有彼得·巴特利特,以及以撰写人工智能教科书而广为人知的斯图尔特·罗素。三人都是各自领域首屈一指的学者,由此可以推知吴恩达的博士阶段经历了怎样水准的学术检验。吴恩达选择乔丹为导师的具体经过,以及两人之间学术交流的轶事,均未流传下来。

博士学位论文的题目是《强化学习中的奖励塑形与策略搜索》。这篇论文在马尔可夫决策过程与部分可观察马尔可夫决策过程——即描述智能体如何在不确定环境中学习最优行动的数学框架——之内,深入探讨了两个问题。第一个是奖励塑形问题。在强化学习中,智能体通过获得奖励来学习,但如果只在达成目标时才给予奖励,学习过程可能会过于缓慢。因此曾有人尝试在中间阶段也加入小额奖励以加快学习速度,但这种做法有可能让智能体找到意想不到的捷径,偏离原本的目标。吴恩达通过数学分析证明,只有一种被称为“基于势函数”的特定形式的奖励函数,才不会扰乱原本追求的最优策略。第二个是策略搜索问题。在环境复杂、状态变化难以精确预知的情况下,与其逐一精确计算每种情形的价值,不如通过运行模拟器、直接试验多种行动方针来得更为现实。为此,吴恩达提出了一种名为“轨迹树”的方法。为验证这一理论,他利用5×5和8×8大小的网格世界环境,以及模拟骑自行车的仿真器等进行了实验。论文所载内容并未止步于抽象公式,而是经由实际运行的系统得到了验证。

关于博士学位的取得时间,不同资料之间存在差异。一份文献记载他于2002年获得学位,但论文原文封面上标注的提交时间却是2003年秋。究竟哪一个才是行政上最终确定的日期,仅凭现存资料尚不足以判断。本章将两种记录原样并列,不断定其中哪一个正确。

在博士阶段即将结束的2002年,吴恩达加入了斯坦福大学教师队伍。他同时担任计算机科学系的兼职教授和电气工程系的副教授。在两个系同时任职这一事实说明,他所研究的课题并未只停留在纯理论或工程应用的某一端。在斯坦福立足之后,吴恩达随即出任斯坦福人工智能实验室(通常简称SAIL)的主任。SAIL发展成一个拥有二十余名教授及多个研究小组的大型学术机构,吴恩达在领导这一机构的过程中,近距离见证了人工智能这一学科在学术上不断拓宽加深的历程。

吴恩达在斯坦福取得的成果中,格外突出的是自主飞行直升机研究。直升机本身是一种极不稳定的机械,只要平衡稍有偏差,旋翼的运动就可能把整个机身带向无法预测的方向,因此即便是人工操控,也需要长期训练。让电脑自主控制这样的机器,在当时属于机器人学与控制理论领域数一数二的难题。吴恩达与他的学生皮特·阿贝尔、亚当·科茨等人一同挑战了这一课题。他们采用的方法被称为学徒学习与逆强化学习——不是由人一条条编写最优规则,而是通过观察熟练的专业飞行员亲自演示的飞行,反过来推断出什么才是好的飞行方式。研究团队使用了一架名为雅马哈R-50的遥控直升机。他们让专业飞行员演示大约十次想要的特技飞行,再从中挑选出最平稳流畅的五段飞行轨迹输入系统。基于这些数据,直升机自主学习出奖励函数与飞行动力学模型,最终实现了倒飞、极限特技飞行、发动机关闭状态下依靠自转下落并自主着陆等媲美专业飞行员水准的飞行。这项研究成为强化学习这一理论框架在真实物理机器上同样奏效的一个例证。

这一时期,吴恩达发表了两百多篇论文,在学界站稳了脚跟。在贝尔实验室涉猎的强化学习课题,经由伯克利的博士论文,延续到斯坦福的直升机项目,这一层层夯实的学术根基,后来成为他将要走上的多条分岔路的起点。此后,他在2011年至2012年间出任谷歌大脑项目的创始负责人;以他在斯坦福开设的在线机器学习课程为起点,共同创办了在线教育平台Coursera;此后又前往中国的百度出任首席科学家。这每一个选择,都将在后续章节中再作详述。不过本章需要明确的是,所有这些选择的背后,都铺垫着一段长达二十六年的时光——从伦敦出生,在香港和新加坡成长,历经卡内基梅隆、MIT、伯克利,最终抵达斯坦福。关于这段时期,几乎没有流传下来任何个人情感或私密故事。留下来的,只有学位、论文、实验室、实验结果这些名称构成的痕迹。然而仅凭这些痕迹,已足以清晰地显示出一个人经由怎样的路径,成长为一位世界级的人工智能研究者。

一个从香港迁往伦敦的移民家庭中诞生了孩子,这在当时的英国社会是常见的景象之一。二十世纪后半叶,英国接纳了相当数量来自昔日殖民地香港的移民,其中不少人以求学或求职为目的在英国定居下来。吴恩达的父母很可能也是在这股潮流中在伦敦落脚,但他们具体出于什么原因决定移民、在伦敦过着怎样的生活,现存资料都无法确认。不过,他被赋予香港裔英国人身份这一事实本身,与他日后成为美国公民、被重新称为香港裔美国人的过程相互呼应,构成了一条完整的脉络。一个人出生的国家、成长的国家、以及成年后取得国籍的国家各不相同,这一事实似乎也与他日后往返美国和中国工作的经历不无关联。虽然没有根据可以断定跨越国境的成长背景直接造就了国际化的视野,但至少可以合理推测,他成年后在多个国家的学校与企业之间往返时,不曾感到特别的心理障碍。

新加坡的莱佛士书院,不是一所用“知名高中”就能简单概括的机构。这所学校由斯坦福·莱佛士于19世纪初创立,此后一直是新加坡教育体制的骨干,即便在独立之后,也始终处于新加坡政府尽早选拔优秀人才并加以集中培养这一政策的核心位置。毕业于这样一所学校,说明吴恩达在十几岁时就已经被归入新加坡教育体系中的顶尖学生之列。新加坡在国际学业评估中的数学与科学成绩一直名列前茅,而这个国家顶尖学校毕业的学生进入美国名牌理工大学,在当时也是并不罕见的路径。不过,吴恩达本人在这所学校具体在哪些科目上表现突出、出于什么契机选择了卡内基梅隆这所特定的大学,都未能确认。因此,这一时期的升学决定,只以成绩单或入学记录之类的行政痕迹留存下来,从未由他本人的口吻加以说明。

再仔细审视他在卡内基梅隆选择的三专业组合,便能看出这并非轻松拿学分的选择。在美国大学,双专业并不少见,但要同时修读分属不同学科领域的三个专业,必须消化每个系各自要求的必修课程。计算机科学系要求算法、数据结构、程序语言理论一类的课程,统计学系要求概率论与推断理论,经济学系则要求微观经济学与计量经济学方法论。这三组课程之间虽有交集,却各自要求不同的思维训练,因此对学生而言,每学期制定选课计划本身就要耗费相当大的精力。吴恩达在承担这一负担的同时还以全班第一名毕业,这一确凿的结果表明,他从本科时期起就已经擅长在多个领域之间自由穿梭。这种能力此后也在他的研究生涯中反复显现。机器学习这一领域本身正处于数学、统计学与计算机工程交汇之处,而吴恩达日后处理的问题,也常常游走于纯理论与实际应用之间。

关于博士导师迈克尔·艾·乔丹的地位,可以再作说明:在统计学与机器学习两个学科几乎互不对话的年代,他被公认为建立起连接两者的理论框架的学者。从乔丹实验室走出的学生中,有多人后来在各自的大学或企业中登上引领机器学习研究的位置,在这样的背景下,乔丹的实验室在当时的伯克利内部也以竞争格外激烈而闻名。列席吴恩达博士学位答辩委员会的彼得·巴特利特,是以统计学习理论中精细的数学分析而知名的学者;斯图尔特·罗素则是系统整理了整个人工智能学科的教科书的作者,几乎每一个初入

这一领域的学生都要通过他的书来认识人工智能,可见其影响力之大。这三人共同评审一篇博士论文,这一事实足以让人推知吴恩达的研究所要经受的检验达到了怎样的水准。学位论文答辩通常是在一位导师和所在学系内外若干位教授组成的委员会面前进行的,由于每位委员都会从不同角度挑剔论文的漏洞,委员会成员本身的分量,也就体现了那篇论文所必须承受的学术重量。

吴恩达完成博士学位后在斯坦福获得的两个职衔——计算机科学系兼职教授和电气工程系副教授——意味着他并未完全归属于某一个系,而是处在可以往返两个系开展研究的位置。在美国的研究型大学,这种双重任职并不少见,但由于所属院系不同,科研经费体系和指导学生的方式也会随之不同,一位新任教授能够同时在两个系获得职位,本身就说明两个系都认为需要他的研究。在计算机科学系看来,吴恩达是研究人工智能与机器学习这一当时重要性已日益凸显的领域的学者;在电气工程系看来,他则是精通信号处理与控制理论、能够为机器人学或自动控制相关问题做出贡献的研究者。正是在这两个系的需求交汇之处,吴恩达不断拓展着自己的研究领域。

再具体审视他领导SAIL实验室时期的规模,拥有二十余名教授这一事实说明,这个实验室已经超越了单一实验室的层级,成为一个学术共同体。年纪轻轻就能领导斯坦福内部数一数二规模的研究机构,这本身也反证了他在学界已经获得了相当程度的信任。实验室主任这一职位,不仅要继续推进自己的研究,还必须协调所属各研究小组朝着不同方向发展的步调,安排新加入的研究生和博士后研究员,并向外界说明整个实验室的方向,承担这一系列职责。吴恩达在担任这一角色期间经历过怎样的行政负担或组织运营方面的具体状况,并没有资料留存,但从他此后新建谷歌大脑这一组织、共同创办Coursera公司、并在百度领导大规模组织的履历来看,可以推测在SAIL的这段经历,正是培养他统筹人力与资源之感觉的地方。

两百多篇论文这一数字本身,就是衡量一位研究者产出能力的指标。一篇学术论文的诞生,通常要经过界定问题、审阅相关的既有研究、设计实验、分析结果、通过同行评审等一系列过程。这一过程被重复了两百多次,说明吴恩达自在斯坦福立足以来,始终同时推进着多条研究脉络。他所涉猎的课题不止于强化学习和机器人学,此后还延伸到计算机视觉、自然语言处理、大规模神经网络训练等领域,这种广泛的研究履历表明,他并未被某一种特定技术束缚,而是持续跟随人工智能这一学科整体的变化,不断调整自己的研究方向。正如直升机研究所展示的那样,他是一位格外注重验证理论能否在真实机器上运作的研究者,这种态度不仅体现在论文数量上,也体现在他所打造的真正可运行系统的数量上。

把这一时期吴恩达的履历整体放在一起看,从一个出生于伦敦的孩子,经由香港和新加坡,穿过美国三所大学,最终在斯坦福立足,所耗费的时间还不到三十年。在这短短的时间里,他先后在不同的大学分别取得学士、硕士、博士学位,曾在世界级的通信实验室担任研究员,后来成为名校教授,并出任大型实验室的主任。推动这一轨迹的个人动机或心理动力,文献中并未确认,但仅凭现存的学位记录、论文与实验室运营履历,已足以充分显示出他在学术世界中拓展自身位置的速度之快、范围之广。截至这一时期,他所建立起来的,还不是广为大众所知的名声,而是在同行研究者之间赢得信任的学者地位。这份信任,很快便把他召唤到实验室之外更广阔的舞台上。

卡内基梅隆大学后来之所以被人工智能研究者称为“摇篮”,自有其历史渊源。1965年,该校在美国大学中率先设立了独立的计算机科学系,核心人物是赫伯特・西蒙和艾伦・纽厄尔两位学者。二人在“电脑能够模拟人类思考过程”这一假设之下,奠定了早期人工智能研究的骨架,并于1975年共同获得计算机科学领域最高荣誉图灵奖。在这两位学者创立的学系这一背景之上,卡内基梅隆此后数十年间始终是机器人学与人工智能研究的中心地之一,到吴恩达入学的1990年代初,这里已经是吸引世界各地优秀学生蜂拥而

至的学校。他之所以能够选择三专业,背后也有这所学校的条件支撑。在拥有涵盖计算机科学、统计学、经济学的广泛课程体系的大学并不多见的年代,卡内基梅隆早已具备足以支撑跨学科教育的规模与体系。

他在MIT时期亲手打造的论文检索系统所处的位置,如果结合当时学界的状况来看会更清晰。1990年代后期,互联网正迅速普及,但学术论文仍然分散在各个学会和期刊各自的网站上,由各方独立管理。研究者要寻找相关论文,往往只能逐一翻查各个网站,或者向同行请教。在这种情况下,能够自动扫描网络、收集并索引论文的系统十分罕见,而一个刚本科毕业的学生亲自设计并实现了这样的系统,这一事实说明他所处理的技术并未止步于理论。这一系统后来演变而成的CiteSeer,此后被多所大学和研究机构用于追踪论文引用关系,长期作为学界的基本工具之一而留存下来。如今,谷歌学术或Semantic Scholar一类的服务已经承担起这一角色,但在这些服务出现之前,学界曾长期高度依赖吴恩达所打造的这类早期系统。

吴恩达在伯克利攻读博士的这段时期,也正是人工智能这一学科本身转变方向的时候。1970年代和1980年代,人工智能研究接连出现不如预期的成果,经历了研究经费与关注度大幅萎缩的时期。学界通常把这段时期称为“人工智能的寒冬”。原因在于,人们发现由人一条条编写规则和逻辑的方式,难以应对现实世界中的复杂问题。经历这个寒冬之后,研究者们不再直接编写规则,转而从数据中寻找统计规律。概率、统计与优化理论,也正是从这时起开始成为人工智能研究的语言。吴恩达的导师迈克尔·艾·乔丹,正是在这一转折时期从事连接统计学语言与人工智能问题这项工作的学者。吴恩达在博士阶段研究的强化学习,同样不是基于规则的方法,而是建立在试错与奖励这一统计的、概率的框架之上,由此可见,他的研究课题本身正处在这一转折的潮流之上。

斯坦福人工智能实验室SAIL,若追溯其历史,同样可以让人掂量出吴恩达所担任职位的分量。这个实验室由约翰·麦卡锡于1962年创立,是“人工智能”这一术语的创造者之一亲手建立的世界最早的人工智能研究机构之一。SAIL早期以机械臂物体操作、自主移动机器人一类的实验闻名,此后也持续在计算机视觉、自然语言处理、机器人学等多个方向上不断产出研究成果。领导这样一个有着深厚履历的实验室,说明吴恩达不仅仅是作为个人研究者获得认可,更是在人工智能这一学术共同体中登上了代表整个机构的位置。

吴恩达获聘为斯坦福教授的2002年,也正是整个硅谷动荡不安的时期。2000年初达到顶峰的互联网热潮,从次年起急剧退潮,大批互联网公司相继倒闭或大幅缩减规模。围绕斯坦福的风险资本与创业生态当时也处于萎缩状态。在这种局面下,大学这一空间相对而言提供了较为稳定的研究基础,吴恩达得以在与产业界的起伏保持一定距离的情况下,在SAIL持续积累学术成果。不过,这一时期硅谷所经历的萎缩并未持续太久。2000年代中期以后,随着搜索和在线服务企业重新增长,硅谷重新焕发活力,大学实验室中锤炼出的技术流向产业现场的通道也逐渐拓宽。吴恩达此后与谷歌这家企业结缘、在校园之外的产业现场检验自己的研究,其背后的时代背景,正是硅谷的这股潮流。

# 早期研究者：机器人与强化学习

直升机在前。那时YouTube视频里的猫和吸引全球学员涌入的在线课堂都还没有出现，吴恩达正在教一架机器人直升机学会在空中自己翻身。他在斯坦福大学任教后最先让世人知晓其名字的领域，不是庞大的神经网络，而是强化学习。这是一种机器人或软件智能体在事先没有得到正确答案的情况下，通过与环境碰撞、反复试错来自行找到更好行为的学习方式。这一时期的研究远离华丽的发布会和大众关注，却是让他培养出处理大规模数据的直觉的时期。

吴恩达在加州大学伯克利分校跟随机器学习领域的世界级学者迈克尔·I·乔丹攻读博士学位。论文答辩委员会中还有彼得·巴特利特和斯图尔特·罗素。他的博士论文研究的是马尔可夫决策过程（简称MDP）这一数学框架，以及它的扩展形式——部分可观察马尔可夫决策过程（POMDP）。这两个概念至今仍是强化学习理论的骨架。如果智能体能够完全知道自己当前所处的状态，就称为MDP；如果只能推测其中一部分，就称为POMDP。机器人仅凭一台摄像头观察世界并做出判断的实际情况，大多更接近后者。

博士论文的贡献之一是奖励塑造这一主题。想让学习速度更快的研究者常常会忍不住在智能体靠近目标点时随意追加一些小额奖励。问题在于，这样被人为调整过的奖励可能诱使智能体找到与原本目标不同的取巧办法。举例来说，如果每当机器人靠近目标就给它加分，机器人可能不会径直朝目标前进，反而会在目标附近打转，靠不断刷分来完成学习。吴恩达从数学上深入剖析了这个问题，证明只有以一种被称为基于势函数的奖励塑造函数的特定形式来调整奖励，才能保证原本的最优策略维持不变。这看似只是一条局部性的定理，实际上却成了判断研究者为提高学习速度而随意设计的奖励何时安全、何时危险的标准。

另一项贡献是轨迹树和Pegasus算法。在复杂的POMDP环境中，精确计算所有状态与行动的价值往往在现实中变得不可行。他没有去完美求解价值函数，而是设计了一种方法：多次运行模拟器得到的轨迹以树状结构串联起来，以此估计策略的表现。Pegasus算法在此基础上更进一步，把带有随机波动的模拟问题转化为固定随机数的确定性形式来处理。如果每次模拟结果都不同，就很难比较出哪个策略真正更优；而固定随机数后，各策略就能在相同条件下得到公平的比较。这套方法论先在5x5、8x8、10x10、50x50等简单的网格世界实验环境中得到验证，但其背后早已埋下了处理远比这复杂得多的物理系统的蓝图。

转到斯坦福任教后，吴恩达的研究团队投入到把这套理论移植到真实机器上的工作中。他们与彼得·阿比尔、亚当·科茨等同事一起选定的对象是自主飞行的直升机。直升机飞行在控制理论中素来被视为数一数二的难题。机身的运动是高维的、不对称的、噪声大、非线性的，还带有非最小相位动力学特性——推动操纵杆的方向与机身实际反应的方向会瞬间出现偏差。此前也曾有人尝试直接手工编写目标轨迹让直升机做特技飞行，但大多数在精细机动动作面前都露出了局限。

研究团队选择的路径是学徒学习：不由人手工编写轨迹，而是让机器人观摩熟练飞行员的示范飞行并从中学习。他们先收集了20分钟的飞行数据，用线性回归建立起基础的动力学模型，这个模型的作用是在给定操纵杆输入的情况下，预测机身大致会如何运动。随后，人类专家把同一个特技动作反复演示了十次左右。即便是人，也无法每次都画出完全相同的轨迹。研究团队使用被称为EM算法的统计方法，

从多次示范飞行中滤除飞行员的失误和抖动，反向推算出隐藏在背后的理想轨迹。这并非原样照搬人的笨拙动作，而是用数学方法还原出人原本想要做出的完美动作。

为了把这样得到的目标轨迹落实到真实飞行中，研究团队采用了每秒运行20次的后退时域DDP（差分动态规划）与高斯-牛顿LQR控制器。实验使用的是名为XCell Tempest和Synergy N9的遥控直升机机体。为精确测量机身的姿态和朝向，他们加装了MicroStrain 3DM-GX1传感器，并在地面架设多台视觉摄像头，以三角测量方式追踪机体位置。自主飞行中新获得的数据会再次反馈到模型中，用来针对特定轨迹对动力学模型进行局部微调。这一过程反复进行多次后，模型和控制器变得越来越贴近真实机体的运动。

结果，这架直升机成为世界上第一架自主完成多种高难度特技飞行的机器：原地翻身的翻转（Flip）、侧向转动的横滚（Roll）、画圆的回环（Loop）、机尾朝下像钟摆一样前后摇摆机身的Tic-Toc、高速倒退画出锥形轨迹的飓风（Hurricane），以及连续反复垂直翻转的混沌（Chaos）。其中Tic-Toc和混沌是靠手工编写轨迹的方式几乎无法成功完成的动作。自主飞行的精度甚至超过了熟练的人类飞行员。在自主飞行表演中，最初的位置误差为3.29米，在把自主飞行中新获得的数据不断整合进模型后，误差缩小到了1.75米。机体自主飞行中获得的经验反过来让机体本身变得更精确，这样一个循环由此确立了下来。

这项研究中最危险的部分，是在发动机完全停止的状态下尝试自转下滑（autorotation）着陆。直升机发动机熄火后，旋翼不再获得动力，但可以利用下降过程中产生的气流让旋翼继续转动，并在着陆的最后一刻利用这份旋转能量。这是即便对真人飞行员而言也需要经过训练的危险机动。研究团队将下降过程中的旋翼转速设定为1150转/分钟，低于正常动力飞行时的1700转/分钟，并设计使前进速度保持在每秒8米、下降速度保持在每秒5米。在距地面9米高度时开始抬起机头、降低速度的拉飘（flare）动作，到距地面0.5米时把速度控制到接近于零，这是一套精密的控制。团队以这种方式共尝试了25次自转下滑着陆，全部成功，机体没有出现一次损伤。

如果说直升机特技飞行是在极端环境中检验强化学习与学徒学习理论的实验，那么同一时期，研究团队也把手伸向了与人协同工作的室内机器人。他们用串联弹性设计和步进电机制造出可替代昂贵工业机械臂的低成本机械臂，其有效载荷为2公斤，最高速度为每秒1.5米，重复定位精度在3毫米以内。用这条机械臂，他们甚至成功演示了在不伤到人的前提下安全地倒出和翻转煎饼面糊的烹饪操作。在斯坦福人工智能机器人项目（简称STAIR）中，团队用配备7自由度机械臂和Barrett机械手的机器人实现了自主收银员：这台机器人仅凭三维传感器拍摄的一张深度图像，就能用手抓住从未见过的物体，一边转动物体一边寻找条形码，并读出其中的数字编码。

围绕提升抓取物体时指尖感知能力的研究也在同步推进。团队自行研制了可安装在Barrett机械手指尖的小型光学近程传感器，在手接触物体之前就能预先感知物体表面朝向哪个方向，并据此建立起能实时修正抓取点的控制器。此外，出于机器人要在室内空间自由移动就必须能自己乘坐电梯这一问题意识，团队还开展了识别并操作电梯按钮面板的研究。针对四足行走机器人LittleDog，团队进行的研究是仅凭立体视觉就能立体地把握眼前的地形，让机器人越过凹凸不平、形状不规则的障碍地形。直升机、机械臂、收银机器人、行走机器人，各自面对的对象各不相同，但其底层始终贯穿着同一个方向：不由人一一编写规则，而是让机器人从数据中自行学习。

这些让机器人的眼睛变得更聪明的研究，自然而然地撞上了同一堵墙。要提升识别物体的能力，最终还是需要人工设计判断所依据的特征，而这种方式有一个局限：每当遇到新的物体或新的环境，就必须重新进行一轮设计工作。研究团队转向了大规模无监督学习——不由人预先设定特征，而是把大量未标注的数据展示给系统，让系统自己找出特征本身。在这一转变中最先碰到的问题是计算速度。当时的深度信念网络和稀疏编码等方法在理论上前景可观，但用一台中央处理器来训练数千万个参数，慢到需要花上几周时间。研究团队开创性地引入了图形处理器（即GPU）来打通这一瓶颈——这个想法是把原本为了在屏幕上作图而设计的硬件，用来训练大规模神经网络。GPU能够同时并行处理大量运算，把同样的训练时间从几周缩短到了以天为单位。

计算速度的问题解决之后，下一个目标转向了神经网络究竟能扩大到什么规模。研究团队提出了一套从海量无标注数据中提取高层次特征的方法论：他们建立了一个包含重建成本、基于TICA（拓扑独立成分分析）的目标函数，用它训练相当于神经网络两层的权重。即使没有人告诉系统要寻找什么物体的正确答案，系统仅仅通过浏览海量数据，就具备了自行识别出人脸或猫头等形状的检测能力。在监督学习中，人必须一张一张地标注这张照片里有猫、那张照片里没有猫；而这个实验表明，即便没有标准答案，仅凭数据的数量也能达到类似的效果。

从强化学习出发，经过直升机的极限控制，再经历处理机器人手和脚的多项实验，最终走到用GPU突破计算速度瓶颈、以无监督学习自行找出特征的地步，这是一条连贯的脉络。每一项研究都是以独立论文的形式发表的，但其背后有一个共同的方向：从由人来制定规则的方式，转向由数据自己生成规则的方式。这个用GPU训练数千万个参数、自行找出人脸和猫头的实验，几年后成了一座桥梁，通向后来那个拥有超过十亿个参数的神经网络在YouTube视频中找出猫的谷歌大脑（Google Brain）项目。作为一名早期研究者，吴恩达在机器人与强化学习领域积累下来的，不是华丽的成果，而是亲身体会到该如何处理数据和算力才能让系统自己变得更聪明的那段时光。

直升机特技飞行之所以在学界引起格外的关注，是有原因的。控制理论研究者很早以前就把直升机当作系统控制的试金石。与飞机不同，直升机即便在低速下也能悬停在空中，还能自由改变方向，但正因如此，旋翼产生的旋转力、反作用力矩和风的影响相互纠缠，使其运动难以预测。由于推动操纵杆的方向与机身实际反应的方向会瞬间出现偏差，传统控制理论所处理的稳定系统模型很难解释特技飞行这类极端机动。此前那些试图靠手工编写轨迹让直升机自主飞行的尝试，最多只能成功做到悬停或平缓的转弯，一旦遇到旋转与反转叠加的高难度动作就屡屡碰壁。正是在这样的背景下，吴恩达研究团队拿出的成果，触及了机器人学与控制理论这两个学科长期未能解决的问题。

构成这项研究骨架的学徒学习概念，也值得再多看一层。在一般的强化学习中，研究者必须事先设定好奖励函数：机器人做出某个行为该给多少分、做出另一个行为该扣多少分，都要由人用数字规定下来，学习才能开始。然而，要把直升机优雅旋转的动作或平顺着陆的手感用一个数字表达出来，比想象中要棘手得多。人类飞行员能凭身体分辨出飞得好和飞得差，但很难把这种判断标准转化为明确的公式。学徒学习反其道而行地解决了这个问题：不由人直接设定奖励函数，而是观察熟练飞行员实际是如何飞行的，反过来推测其背后的目标。即便人的示范飞行并不完美，只要用统计方法从多次尝试中滤出共同存在的理想轨迹，就能把连人自己都说不清楚的那种手感也变成学习的对象——这正是这一思路的出发点。这种方法后来在机器人学、自动驾驶等诸多领域，成为解决那些人难以设计出奖励函数的问题时反复被采用的方法论。

在斯坦福主持这项研究的实验室，以STAIR即斯坦福人工智能机器人项目之名，把多个方向的工作汇聚在了一起。负责直升机飞行的团队、用机械臂抓取物品的团队、打造收银机器人的团队，表面上解决的是各不相同的问题，实际上却是在同一间实验室里彼此分享人员和设备、相互协作的关系。彼得·阿比尔在直升机特技飞行研究中承担了打磨学徒学习理论的角色，此后在斯坦福机器人实验室把研究领域拓展到了抓取与操作问题。亚当·科茨参与直升机控制实验之后，转向了大规模无监督学习研究，并接手了用GPU训练神经网络的实验。当一个人在摆弄机械臂、另一个人在放飞直升机、又一个人在调试收银机器人的条形码识别时，他们彼此分享着各自的成果，共同培养出了一种让系统从数据中学习的态度。可以说，源于迈克尔·乔丹伯克利实验室的数学严谨性，与斯坦福实验室培养出来的硬件实验文化，就这样在同一个人身上重叠了起来。

先在网格世界这种简化的格子环境中验证理论，然后才把它移植到真正的直升机这种昂贵而危险的机器上——这样的顺序自有其道理。网格世界是一个虚拟的棋盘状空间，机器人只能上下左右移动。在这样的小环境里，可以快速而安全地确认算法是否按理论运作、策略搜索是否真的能找出更好的行为。如果直接用真实的直升机开始实验，一旦算法存在缺陷，就可能导致机身坠毁、昂贵设备损毁。只有在5x5、50x50等各种规模的网格中反复验证了Pegasus算法之后，研究团队才把同样的原理扩展到20分钟的飞行数据与EM算法、后退时域DDP控制器上，应用到真实机体。这种把在小问题上验证过的原理谨慎地推广到大问题上的做法，此后在吴恩达处理的许多研究中反复出现，成了一种一贯的态度。

在机械臂抓取物品的研究中，也贯穿着类似的问题意识。收银机器人要抓住从未见过的物品，就必须把摄像头或深度传感器拍到的影像中物品所在的位置和形状，转化为计算机能理解的数字，这些转化出来的数字就被称为特征。物体的棱角在哪里、表面朝哪个方向倾斜、颜色如何分布——这类信息长期以来都是由人类研究者一条条制定成规则、再告诉计算机的方式来处理的标准做法。问题在于，这些规则只对特定的物体或环境管用。一旦收银台上出现新商品，或者灯光发生变化，事先编写好的规则就常常失灵。在指尖加装近程传感器、预先感知表面方向并实时修正抓取点的研究，归根结底也是想用传感器去弥补那些光靠人定规则难以处理的变量。识别电梯按钮面板的研究，或是LittleDog的地形识别研究，也都处在同一个脉络之中。如果机器人每遇到一个新空间、新物体，都需要人重新编写规则，那么很难称得上真正意义上的自主。团队越是想让机器人的眼睛和手变得更聪明，人工设计规则的局限就越是反复暴露出来。

这样的经验，成了研究团队把方向转向让特征不由人来设定、而是从数据中自行学习的契机。基于TICA的目标函数，是以重建成本——即神经网络能把输入的图像还原得多接近原貌——作为标准来训练权重的。即使没有人告诉它这张照片里有脸、那张照片里有猫这样的正确答案，神经网络仅靠浏览海量照片，也能自行捕捉到反复出现的形状。这个原理本身并不是全新的想法，深度信念网络、稀疏编码等方法论早已为学界所知。只是要把这类方法运行到有实用价值的规模，需要反复调整从数千万到数亿个权重，而当时常用的一台中央处理器要完成这样的运算，得花上好几周。因此，研究团队引入图形处理器的这一选择意义重大。这种原本为在屏幕上绘制三维图形而设计的硬件，恰好优化于同时处理大量的小规模运算。神经网络的训练归根结底也是数不清的数字同时相乘相加的重复运算，于是这个为绘图而生的装置，出乎意料地也很适合用来训练神经网络。当计算所需的时间从几周缩短到几天、又从几天缩短到以天甚至更短为单位时，研究团队便得以尝试此前想都不敢想的规模的神经网络。

当时的计算机视觉学界，仍以使用人工设计的特征作为标准做法。仅凭数据、没有正确答案标签就识别出人脸或猫头这类形状的研究，在学界并没有立刻被广泛接受。然而，这项实验展示的结果是明确的：即便没有人一一告诉系统该找什么，只要给足够大的神经网络展示足够多的数据，系统就能自行找出有意义的模式。这个结果在学界内部也受到了怀疑的目光。当时神经网络这一方法论本身正处于被视为过时的时期，而靠增加数据量和计算量来解决问题的做法，在偏好建立精细数学模型的研究者眼中，多少显得是一种粗糙的解法。尽管如此，吴恩达的研究团队还是把从直升机实验中确认的态度——与其由人一一规定正确答案，不如让数据自己找到答案——原封不动地带到了计算机视觉领域。

自转下滑着陆实验所具有的分量，也值得再度审视。发动机熄火状态下仅凭旋翼的旋转能量安全降落到地面的这一机动，对真人飞行员而言也是需要长期训练和胆量的应急程序。研究团队把这种着陆反复尝试了二十五次，且全程没有一次机体损伤，这一事实表明，学习出来的策略不仅在平稳条件下管用，在失去动力的极端情况下也能始终如一地运作。从旋翼转速、下降速度，到开始拉飘机动的高度都经过细致设计的这项实验，同时也是在检验从数据中学得的策略是否可靠到足以满足人所设定的安全标准。要把学习出来的系统投放到实验室之外的危险情境中，必须经过这样反复的验证——这种直觉，此后在处理机器人与自主系统的诸多研究中，成了共同遵循的标准。

回过头看，这一时期的研究虽然分别面对不同的机器和不同的问题，却拥有共同的实验步骤：先用数学方式将问题形式化，在小而安全的环境中验证这一形式化是否正确，然后才把它移植到真实硬件上确认结果。在直升机特技飞行中，从网格世界出发走向真实机体；在机械臂研究中，用传感器和数据填补人定规则的局限；在无监督学习研究中，用图形处理器打通计算速度的瓶颈。这三个方向的工作以不同的论文形式发表，但实际上是在斯坦福同一间实验室里、由同一批人往返其间共同积累起来的一种实验文化。不由人事先规定正确答案、而是让数据自己找出规则——这种态度，成了自然而然通向下一阶段工作的一条通道，而那个阶段将要处理规模远大得多的神经网络和远为庞大的数据。

与直升机那种如山脊般起伏蜿蜒的运动不同，强化学习理论大体上有两条主要路径。一条是为每个状态和行动都标定一个数值、选取数值最大的一方的价值函数估计方式；另一条是不单独计算数值，而是直接改变策略本身、寻找性能提升方向的策略搜索方式。Pegasus算法选择的正是后者。在复杂的POMDP中，精确标定每一个状态的数值这件事，本身往往因计算量过大而难以承受。于是研究团队没有去计算数值，而是选择把策略在模拟器中多次运行，比较得到的各条轨迹，从中挑出更好的策略。基于势函数的奖励塑造这一概念，也与这一选择相互呼应。这里所说的势，其思路类似物理学中处于高处的物体所具有的位能。如果事先为某个状态标定一个表示其与目标有多接近的数值，并且机器人每次转移状态时，只按这个数值的差额来增加或扣除奖励，那么无论机器人绕多少弯路，原本最优路径的排序最终都不会被打乱。这与只要起点和终点的高度差相同、无论沿哪条路走上坡最终获得的位能都一样这一算法并无二致。

在机械臂抓取物品的研究中，还有一点值得留意。与收银机器人研究并行开展的工作中，有一项是学习多个手指同时接触物体时形成的多个接触点。人的手要稳稳握住一个杯子或瓶子时，光靠一根手指接触是不够的，必须有两个以上的位置从不同方向按压物体，物体才不会滑动或转动，从而被稳稳握住。研究团队把这一原理搬到机器人手上时，尝试的方法不是由人一一制定哪根手指该放在哪里才稳的规则，而是从物体的图像数据中学习出多个接触点的组合。由于每个物体的形状各不相同，能够抓取的点的组合数量也几乎无穷，人靠手工制定规则、事先应对所有可能的情况原本就是不可能的。训练出来的模

型即便面对从未见过的物体，也能从其表面形状中自行挑选出能够稳定握持的接触点组合。收银机器人要找到条形码，就必须在不让物体从手中滑落的前提下把它转来转去，因此这项能够稳定抓取接触点的学习，几乎是整个收银机器人得以运作的基础。

STAIR这个项目名称的由来也自有用意。直升机团队、机械臂团队、收银员团队、电梯识别团队、行走机器人团队各自以不同的论文发表成果，但都被归拢在STAIR这一把伞下面。这是因为他们设定了这样一个目标：一台机器人应当能够自己在房间里四处走动、拾取物品、乘坐电梯前往其他楼层，并找到从未见过的物品上的条形码并读取出来——所有这些都要能独自完成。无论是让LittleDog行走室内地形的腿，还是收银机器人的手，抑或按电梯按钮的手臂，这些能力归根结底都是一台机器人要在人类居住的空间里自主行动所需要的一个个组成部分。不同的团队用不同的硬件做实验，却共享着从数据中学习的方法论，原因就在于他们最初要解决的问题原本就是一个连成一体的目标。

贯穿这一时期研究的态度，是在处理大问题之前先用小问题验证这一原则。这也正是Pegasus算法要在从5x5到50x50等各种规模的网格中反复测试的原因。网格每边的长度每增加一分，机器人可能处于的状态数就会以平方的速度膨胀。只有先确认在小网格上行得通的方法在规模大得多的网格上是否依然站得住脚，研究团队才能把这一原理移植到直升机这种昂贵而危险的机器上。这一顺序，在转向机械臂接触点学习或大规模无监督学习时也同样重复出现：先在小问题上确立原理，确认这个原理在问题规模扩大后依然站得住脚，然后才推进到下一个规模——这一顺序，几年后在同时扩大神经网络规模与数据量的工作中，依然延续成了一种一贯的态度。这个逐一处理机器人的手、脚、眼睛的时期，归根结底也是一段小心翼翼地扩大问题规模、亲身确认其极限所在的训练历程。

# 谷歌大脑与规模化实验

2011年的谷歌内部有一个奇怪的团队。它既不是搜索或广告那样支撑公司营收的部门,也不是人们通常想象的那种只发论文的研究机构。这是一个把公司资源押在神经网络这一当时尚未进入学界主流的技术上的团队。这个团队的名字叫谷歌大脑,创立并领导它的人是吴恩达。

故事要追溯到2010年。时任斯坦福大学教授的吴恩达向谷歌管理层递交了一份大胆的提案,内容是调动庞大的计算资源和海量数据来训练神经网络。谷歌接受了这一提案。次年,也就是2011年,吴恩达创立了深度学习项目谷歌大脑,并出任总负责人。

在当时的人工智能学界,专注于扩大模型规模的做法是一种陌生的思路。神经网络研究长期以来被视为比拼算法精巧程度的领域,而堆砌大量计算机来解决问题的想法常常被看作不够优雅的做法。一些同行研究者甚至劝吴恩达不要沉迷于大规模神经网络和规模扩张,认为这对研究者的职业生涯没有帮助。然而,吴恩达赋予谷歌大脑团队的第一个也是唯一的目标十分明确:把神经网络做大,并向它灌注数据。

这份信念从何而来?其根基在于从人脑运作方式中得出的一个假设,即所谓的单一学习算法假说。在创立谷歌大脑之前,吴恩达为了从人脑运作原理中找到打造更好AI算法的线索,曾大量阅读神经科学论文,并与神经科学家交流。但在确认当时的神经科学几乎尚未揭示大脑运作原理之后,他放弃了直接将神经科学应用于智能实现的路径。尽管如此,生物大脑所具有的通用性,依然深深留在了他的思考中。

这种通用性的含义是这样的。人类的DNA长度有限,能承载的信息量也因此受限。然而人类却能完成从攻读数学博士学位到骑摩托车、再到在呼叫中心接听电话等种种截然不同的任务。原因不在于大脑天生就为每项任务配备了专用零件,而在于大脑天生具备一套能在经验和训练中自我专精化的适应能力,一套学习引擎。吴恩达认为人工神经网络也可能如此。他相信,比起针对特定问题的精巧设计,只要有足够的数据和足够的计算资源支撑,单一的神经网络结构也能展现出堪比人脑的通用性。

这份信念很快引出了一个问题:正如婴儿无需他人教导就能自行掌握“脸”这一概念一样,机器是否也能仅凭投喂的数据、而无需任何标准答案,自行掌握世界的种种概念?谷歌大脑团队决定把这个问题变成一场实验,验证仅用完全没有标签的数据,能否训练出对高层次、且对特定对象有选择性反应的特征检测器。这里所说的特征检测器,指的是神经网络内部的神经元,它检测的不是图像中眼睛、鼻子、耳朵这类局部,而是像整张脸或整只猫这样一个鲜明的整体对象。脑科学中把这种只对某一特定对象作出反应的神经元称为“祖母细胞”,这个说法源自这样一个假说:大脑某处存在一个只在看到祖母面孔时才会被激活的神经元。谷歌大脑团队提出的问题是,这种“祖母细胞”式的现象能否仅凭无监督学习、在没有任何指导或奖励的情况下,在人工神经网络中自行出现。

研究团队为回答这个问题,用从YouTube视频中随机抽取的1000万张图像构建了一个数据集。每张图像都是200x200像素的彩色图像。模型是一个由九层组成的局部连接稀疏自编码器,并在此基础上应用了L2池化和局部对比度归一化。自编码器指的是一种神经网络结构,它通过反复压缩输入数据再将其还原的过程,自行学习数据的核心结构。这一神经网络拥有十亿个连接,也就是参数。考虑到当时世界上已报告的大型神经网络大多只有一千万左右的参数,这一规模超出了十倍以上。当然,与人类视觉皮层所拥有的神经元和突触数量相比,这仍然只是其百万分之一的规模,依旧是一个很小的网络。

要训练这样规模的神经网络,需要相应的计算能力来支撑。研究团队动用了由1000台计算机、共1.6万个核心组成的集群,连续运行了三天。为了承担这种规模的计算,研究团队新开发了一个名为DistBelief的软件框架。它把神经元和权重分散存储在169台各配备16个核心的机器上,实现了模型并行化。在此基础上,又结合了数据并行化:将核心模型复制五份,通过256个参数服务器分区,以异步随机梯度下降法进行训练。也就是说,他们把一个神经网络拆分到多台计算机上存放,同时用不同的数据切片训练多个副本,并让这些副本的结果持续相互交换。如今已成为大规模模型训练常识的这套方式,在当时是谷歌大脑亲身摸索出来的全新设计。

结果震动了学界。研究团队用含有人脸的图像和随机混入的干扰图像做分类测试,发现神经网络内部自行形成了一个只对人脸有选择性反应的神经元。这个神经元的准确率为81.7%。更有意思的是接下来的实验:研究团队使用名为OpenCV的图像处理工具,人为地从训练数据中剔除含有人脸的图像,再重新进行同样的实验。结果这个神经元的性能骤降至72.5%,只相当于简单线性滤波器的水平。这一结果表明,神经网络并非偶然获得了这样的神经元,而是真正从数据中自行学会了“脸”这一概念的存在。

这个神经网络的能力并未止步于人脸。自行学会“猫脸”和“人体”概念的神经元也相继出现。在只使用立体黑白图像左侧图像的实验中,检测人体的神经元达到了76.7%的性能,检测猫的神经元达到了74.8%。两项数值都大幅超过了以简单线性滤波器或随机权重为基准的对照组。学到的高层特征还表现出稳定的性质:无论图像左右平移、稍微旋转出画面,还是放大缩小,其反应都不会动摇。

研究团队并未止步于此,他们又在这套学到的特征表示之上叠加了一个逻辑回归分类器,挑战识别ImageNet两万个对象类别的任务。在2000台机器上经过一周的微调后,取得了15.8%的准确率,相较当时的最佳水平实现了70%的相对提升。这项实验后来被世人称为“YouTube猫”实验。一个没有任何标准答案、只看过YouTube视频的神经网络竟自行认出了“猫”这一概念,这个消息不仅在研究者中间,也在普通大众中成为一段有趣的谈资,并成为深度学习这项技术首次进入大众媒体视野的事件之一。

这段时期,吴恩达仍在斯坦福讲台上教书,与此同时,日后问世的在线课程和Coursera的筹备工作也在同一时期并行推进。谷歌大脑的实验室、斯坦福的教室,以及日后将发展为Coursera的筹备工作,在同一几年间齐头并进。

YouTube猫实验证明了大规模深度学习所蕴含的学术潜力,但这并不足以说服整个公司。谷歌内外仍有不少人对深度学习的实际效用心存怀疑。实验室里得出的有趣结果,与把这项技术嵌入能创造营收的实际产品之间,存在着巨大的差距。为了把这项技术带出实验室、变成整个公司的能力,吴恩达制定了自己的一套说服策略。他所选择的原则是:与其瞄准那种能一举扭转公司战略方向、看起来最有价值的项目,不如在最初的几次尝试中就先取得实际的成功。

按照这一原则,他为谷歌大脑团队选定的第一个内部客户是谷歌语音识别团队。当时,语音识别在公司损益中的分量,不及把AI应用于网页搜索或广告、从而大幅拉动公司营收那样重要。但它在公司内部依然是一项明确有意义的任务。谷歌大脑团队与语音识别团队携手合作,用深度学习显著提升了谷歌语音识别的准确率。

这次合作的成功,成为谷歌内部其他团队开始信任谷歌大脑和深度学习的契机。一次成功往往会带来下一次成功。紧接着,谷歌大脑迎来了第二个内部客户——致力于提升地图数据质量的谷歌地图团队。凭借两次接连的成功获得的推动力,吴恩达终于得以与谷歌核心营收部门广告团队展开深入对话。一场始于实

验室角落的实验,在短短几年间走进了公司的核心地带。

从规模适中却确有意义的项目起步,一步步积累成果与信任,最终把一整家庞大的科技公司改造为AI公司,这种方式并未止步于一次成功故事。日后吴恩达在Landing AI撰写的《AI转型手册》中提出的第一条策略——执行试点项目以获取推动力,其根源正在于此。他在谷歌大脑亲身体会到的这套方法,此后成了他每每向众多企业谈论AI落地时反复提起的原则。

吴恩达在谷歌大脑埋下的规模扩张这套逻辑,此后依然延续了下来。这个团队日后发明了历史上可扩展性最强的神经网络结构——Transformer,为今天正在改变世界的生成式AI浪潮奠定了基础。这种把模型做大、向其中灌注数据、让它模仿数据中隐藏模式的方法,此后成为贯穿整个AI产业的一条法则。

然而,吴恩达本人恰恰是当今最坚决对“通用人工智能(AGI)几年内即将到来”这类说法保持警惕的人物之一。他与那些为了宣传或融资而强调AGI迫近的诸多企业言论保持距离。他比大多数人都更长久地相信规模扩张的力量,却也指出,如今最前沿的模型几乎已经读遍了互联网上的文本,因此仅靠沿用旧办法一味扩大规模的路径正遭遇瓶颈。他的判断是,要突破这一瓶颈,就必须承担呈指数增长的成本。

在他看来,距离真正的AGI还需要几十年的时间。他判断,仅凭对当下Transformer结构原样放大的扩展法则,无法抵达真正的智能。他表示,除了无限做大模型之外,还需要同时建立能把结构化的人类知识植入系统的智能体工作流,也需要与现在不同的新技术突破。

作为开启大规模深度学习时代的人,他给出的这份诊断出人意料地冷静。当年为了在YouTube视频中找出一只猫,让1000台计算机连续运转三天的那个人,如今却说存在一堵无论怎样增加计算机也无法跨越的墙。曾经笃信规模的人,如今率先指出了规模的极限,这份清醒,或许正是他在谷歌大脑学到的另一个教训。

要完整理解谷歌大脑的这项实验,有必要先回顾一下当时人工智能研究处理数据的方式。2010年前后的计算机视觉研究,主流是建立在人工逐张为图像打标签的数据集之上的监督学习。图像中是否有猫、有汽车、有人,都由人预先标注好,神经网络则把这些标注当作标准答案,学习如何做出正确判断。要建立像ImageNet这样包含众多类别的数据集,需要动用大量人力,而给每一张图像贴标签,都是耗时耗钱的工作。即便想扩大数据集的规模,也始终摆脱不了人工能够贴出的标签数量这一明确的天花板。

谷歌大脑的YouTube猫实验所提出的问题,恰恰动摇了这一前提本身。问题是:如果不给神经网络任何人工标签,只让它看一大堆从YouTube随机抽取的视频图像,会发生什么?没有标签,意味着没有任何一只手告诉它“这张图是猫”“这张图是人”这样的标准答案。神经网络必须仅凭图像中反复出现的模式作为线索,自行建立起对世界进行分类的标准。这种方式被称为无监督学习。无监督学习究竟能否产生真正有用的结果,在当时仍是一个未经验证的问题,而谷歌大脑用1000台计算机和三天时间,给出了自己的答案。

“祖母细胞”这一说法,在这一脉络下也值得重新审视。神经科学界长期以来一直存在争论:大脑中某一个神经元究竟是只对某个特定对象作出反应,还是由多个神经元共同协作来表达一个概念。如果真的存在一个每次看到祖母面孔就会激活的神经元,那就意味着大脑正以一种惊人高效的方式,存储着世界上一个个复杂的概念。谷歌大脑团队把这一概念原样搬到人工神经网络中进行实验,等于是用计算机科学的语言重新提出了神经科学的一个古老问题。他们想确认的是:即便无法原样复刻人脑,大脑所展现出的这种

高效的概念表达方式,能否在人工神经网络中得到再现。

“稀疏自编码器”这一名称中的“稀疏”二字也值得展开说明。自编码器在压缩输入再将其还原的过程中,有一条约束条件叫作稀疏性:它要求内部大多数神经元无论遇到什么图像都保持沉默,只有极少数神经元作出反应。如果放任所有神经元都活跃地作出反应,神经网络就会把图像中琐碎的噪声也一并记住;而一旦抑制作出反应的神经元数量,神经网络就被迫只聚焦于真正重要的少数几项特征。“局部连接”指的是这样一种设计:神经网络中的每个神经元只观察图像中自己附近的一小片区域,而非整张图像。这背后的想法是,正如人只看照片的一个角落,也能判断那里是不是眼睛或鼻子一样,每个神经元并不需要看到整张图像。正是靠着这种局部连接,才得以在大幅减少连接数量的同时,依然达到十亿参数这一规模。池化是把多个神经元的反应汇总为一个,使图像即便发生轻微移动或抖动也能给出相同答案的装置;局部对比度归一化则是一种即便图像的亮度或对比度因拍摄环境不同而参差不齐,也能让神经网络保持稳定的装置。

这项实验所使用的计算机的性质也值得留意。如今一提到训练大规模神经网络,人们脑海中往往首先浮现的是大量调用图形处理器,也就是GPU的画面。然而谷歌大脑最初的这项实验,运行的并非GPU集群,而是由普通中央处理器,也就是CPU组成的集群。1000台计算机、1.6万个核心这样庞大的数字之所以必须如此庞大,原因之一正在于此。要用中央处理器去承担一台图形处理器就能胜任的并行运算量,就必须动用更多的机器来弥补不足的算力。在深度学习训练中全面采用图形处理器成为业界标准之前,还需要再过几年。谷歌大脑的这项实验,正是发生在这一转变到来之前、靠机器数量硬生生解决规模问题的一个案例。

再进一步展开DistBelief所实现的并行化方式是这样的:把一个庞大的神经网络整体搬到一台计算机上进行训练,从一开始就是不可能的。研究团队把神经网络切割成多个片段,分散存放到169台机器上,这些被拆分的片段必须彼此不断交换信息、如同一个整体那样协同运作。在此之上,他们又制作了这个被拆分的整个神经网络的五份副本,让每份副本同时学习不同的数据切片。这五份副本学到的内容,不断被搬运到256个参数服务器上进行合并。问题在于这一过程并不能严丝合缝地同步运转:有的副本计算得快,有的副本完成得稍晚。之所以选择按到达顺序接收计算结果的异步方式,是因为如果等待所有机器以相同的速度全部完成,整体训练速度就会被最慢的那台机器拖累。这种异步处理方式,此后在处理大规模神经网络的众多实验室中成为常见的设计方式。

要真正运行这种规模的实验,需要付出相当高昂的代价。让1000台计算机连续运转三天,不是一般大学实验室能承担的预算。身处大学的吴恩达之所以能够构想出这种规模的实验,背后正是谷歌这家公司所拥有的庞大数据中心资源。学界的神经网络研究之所以长期停留在几台计算机就能解决的小问题上,原因之一也正是这种资源上的落差。大学实验室没有办法筹措到这种规模的计算资源,但为了运营搜索、广告、社交网络服务而早已建有庞大数据中心的企业,情况则大不相同。可以说,吴恩达离开学界、转投产业界的这一选择,不仅仅是为了获得更多研究经费,更是把自己转移到了一个能够真正把学界此前连尝试都无法尝试的规模实验付诸实施的环境。

与语音识别团队的合作所具有的意义,也值得再深入看一看。语音识别是把人的声音转换成文字的技术。对着智能手机说话,话语随即变成文字显示在屏幕上,这一过程背后正是这项技术在发挥作用。如果这项技术的准确率不高,用户就会每天都遇到屏幕上显示的句子与自己所说的话不一致的情况。反过来,一旦准确率提高,用户往往察觉不到这种变化,只会单纯地享受到便利。谷歌大脑实现了这种不动声色的改

进,这一事实本身,就是在公司内部悄然证明了深度学习即便没有华丽的发布会或新产品上市,也能实实在在地提升现有服务的品质。比起一篇出自实验室的论文,每天无数用户实际体验发生了变化这一事实,恐怕才是真正推动谷歌内部其他团队的力量。

如果放在这样的背景下阅读,吴恩达如今关于规模扩张之局限所提出的诊断,其脉络也会变得更清晰一些。他在谷歌大脑时代所面对的局限,是计算机的数量和收集数据的速度。只要调动更多计算机、从YouTube抓取更多图像,问题就能得到解决。而他现在所指出的局限,性质却不同。互联网上已有的文本是有限的。如果模型已经几乎读遍了这些文本,那么无论再增加多少计算机,也不再有更多新文本可读。他的诊断是:谷歌大脑时代还有YouTube这样一座广阔的新数据仓库,而如今已不再有这样一座唾手可得的新仓库。区别在于,做大规模这件事本身依然是有效的方法,只是可供做大的原料已不像过去那样充裕。

“智能体工作流”这一表述也值得展开说明。它指的不是对一个问题一次性给出答案的方式,而是把一项任务拆分成多个步骤,让模型先制定计划,再按计划执行,自行重新检查自己给出的结果,并在需要时进行修正,如此形成的一整套流程。这与人在解决难题时,不是直接写下答案,而是先写一份草稿,再重新读一遍草稿、修正错误之处的过程颇为相似。吴恩达之所以把这种方式与仅仅做大模型的旧办法放在一起谈论,是因为在他看来,智能并非仅凭一次庞大的计算就能完成,而是要经过多个阶段的检查与修正,才能最终得到值得信赖的结果。

回过头看,谷歌大脑这一个团队留给世界的,并不止于一段“神经网络认出了猫”的故事。这项实验证明了只要增加计算机数量、灌注数据,神经网络就能自行学会世界的种种概念,而这一结果,正是此后十余年间整个人工智能产业工作方式的预告片。如今训练大型语言模型的研究机构们视为理所当然的流程——尽可能多地聚集计算机、尽可能多地获取数据,再交由神经网络自行找出其中的模式——这套顺序,正是在2011年谷歌的一间实验室里第一次成形。吴恩达是那间实验室的开创者这一事实,让他如今谈论规模扩张之局限时的发言,也承载着不同寻常的分量。

谷歌大脑这个名字背后,并非只有吴恩达一个人。在设计并执行YouTube猫实验的研究团队中,有一个名字叫作Quoc V. Le。这位日后将在神经网络结构搜索领域树立独立声誉的研究者,当时正是谷歌大脑的核心成员,负责实验的具体设计。如果说吴恩达负责确定团队方向、说服公司内部,那么搭建九层自编码器、在1000台计算机上运行代码、逐一核对结果这些工作,则是由多位研究者共同分担的。这项成果从一开始就不是一个人的成果,而是整个团队的成果。

“梯度下降法”这一说法也有必要说明一下。训练神经网络,归根结底是不断微调每一个连接上所赋予的数字,也就是权重。当神经网络给出的答案有误,系统就会计算出这份误差,并朝着减小误差的方向,把权重挪动一点点。把这个动作重复数百万次,神经网络就会逐渐给出越来越准确的答案。这一过程与在大雾弥漫的山中,仅凭脚下感受到的坡度、一步步寻找更低处走下山的样子相似,因此得名“梯度下降法”。之所以加上“随机”二字,是因为它不是查看全部数据后再决定方向,而是随机抽取一部分数据来估算方向。若每次都要核对全部数据,耗时会过长,因此频繁地按估算出的方向小步移动,实际上反而能更快抵达目的地。“异步”一词的意思是,让多台计算机各自按自己的速度不断进行这种估算和迈步,彼此互不等待。

ImageNet实验中出现的“微调”和“逻辑回归分类器”这两个说法也值得展开说明。所谓微调,指的不是把已经完成过一次训练的神经网络整体推倒重来,而是在它已经学到的内容基础上,针对新任务再稍

作调整打磨。当把已经从YouTube视频中学会“脸”和“猫”这些概念的神经网络,拿去用于识别ImageNet两万个类别这一新任务时,团队选择的不是从头重新教起,而是在已有知识之上,只增添适应新任务所需的感觉。逻辑回归分类器则是接收神经网络提取出的特征、对这张图像属于哪个类别作出最终判断的最后一道装置。神经网络的大部分负责从图像中提取有意义的特征,而位于其上的这个小小分类器,则根据这些特征贴上标签。

选择语音识别团队而非广告团队作为第一个内部客户,这一决定也体现出对组织的把握。若把新技术直接推给支撑公司营收的核心部门,一旦失败,这份失败很可能反噬到人们对神经网络这项技术整体的信任。相反,如果先在公司内部虽有意义、却不直接关系营收的部门取得成果,失败了负担也不大,成功了则能成为拉动下一次合作的依据。信任并非一次就能大量获得,而要靠多次小小的成功积累,才能最终打开更大部门的大门,这一点是吴恩达在谷歌内部亲身体会到的。

当时围绕神经网络的怀疑论也有其背景。人工智能研究在上世纪70年代和80年代先后经历过两次所谓的低潮期。每一次,神经网络都未能拿出预期中的成果,研究经费和关注度随之中断。直到2000年代后期,在计算机视觉或语音识别领域,支持向量机或决策树这类方法仍被认为能给出更稳定、更可预测的结果,而神经网络则常常被视为理论上有趣、实战中却难以驾驭的技术。越是了解这段历史的研究者,对再次重仓押注神经网络就越发谨慎。谷歌大脑的成果之所以在学界内外引发巨大震动,原因之一正在于:一项此前多次令人失望的技术,这一次用真实的数据和真实的数字,证明了这次不一样。

这一时期的吴恩达,虽然已经凭一项足以让自己名字留在论文史上的成果崭露头角,却把更多心力投入到让这项成果在公司内部扎根这件事上。证明一项技术的价值,与让整个组织接受这项技术,是两种截然不同的努力。在谷歌大脑,他把这两种努力都做到了,而这段经历也成了他此后往返于学界与产业界之间反复面对的课题——让新技术真正被人使用——的原型。

# 在百度的三年

2014年5月,这位穿梭于斯坦福课堂和Coursera服务器之间、逐渐声名鹊起的学者,迎来了一个新的选择。主导中国互联网搜索市场的企业百度宣布,将在加利福尼亚州森尼韦尔设立美国研发基地,并任命他为首席科学家。这一消息在硅谷和学术界都引起了关注。作为谷歌大脑的创始负责人,他这次转向的不是一家美国企业,而是一家中国企业的研究组织,这一事实本身就颇为罕见。

他在做出这一决定的同时,辞去了Coursera联席首席执行官的职务,但仍保留了董事会主席的头衔。这并非与自己创立的在线教育平台彻底断绝关系,而是从运营第一线后退一步,转向监督者的位置。他放弃硅谷的稳定根基和自己一手打造的Coursera经营职务,转而选择中国企业的内在原因,现存资料无法确认。本章所涉及的事实,是依据百度开业发表文、他辞职时亲笔撰写的文章以及此后广播采访中的发言重新构建而成的。由于公司内部的经营指标或非公开会议记录并未留存,因此本章不涉及他的个人动机或在中国当地生活中可能经历的文化体验。

执掌百度的董事长李彦宏在谈及他的加盟时,称他是真正具有远见的人物,也是人工智能领域的核心贡献者,并表示在AI作用日益扩大的时代,找到了能够引领研究的合适人选。这一评价并未止步于一句寒暄客套话。后来吴恩达回忆李彦宏时,称他是少数几位早早看出深度学习价值的大企业首席执行官之一,直到如今仍是世界屈指可数的AI经营者。两人的关系似乎超越了一纸简单的雇佣合同,发展成彼此信任对方技术判断的合作关系。

他所领导的百度研究院在这一时期被重组为三个研究所。新设立的硅谷AI实验室、自2013年起便持续开展深度学习研究的北京深度学习研究院,以及北京大数据研究院被统合在同一个体系之下。就任之际,他将百度评价为一家拥有长远眼光和深厚投入的公司,并表示很期待能参与到能够改变世界的基础技术的发展中来。这番发言背后隐藏的个人野心或学术上的盘算,并未留存于记录之中。不过,他在这一时期做出的诸多判断此后确实转化为了实际的组织运营,这一点可以从此后数年的行动轨迹中得到印证。

据他本人回忆,2014年至2017年在百度任职期间,他统领着包括约300人规模的百度研究院在内、总人数约达1300人的整个百度AI团队。这一数字并非来自公司官方统计资料,而是出自他本人的回忆,因此难以核实准确的组织架构或人力配置。但从种种迹象来看,规模本身确实不小。至于他实际上是如何管理这样一个庞大组织的、设置了怎样的考核体系、将硅谷式研究文化移植到中国本地组织的过程中出现过哪些摩擦,现存资料均无法确认。这一部分留作记录的空白,才是诚实的叙述方式。

尽管如此,他辞职时留下的文章中提到了多位与他共同领导团队的同事姓名。他与拥有丰富AI经验的经营高管陆奇首席运营官密切合作,也与既是研究者又是技术领袖的王海峰、兼具技术与业务统筹能力的林元庆一同搭建了组织架构。除此之外,他还回忆称自己发掘了包括亚当・科茨在内的多位研究人员,并在夯实组织的知识基础上倾注了心力。这给人留下的印象,与其说是一个聚集在某一个人名下的组织,不如说是一个由多位领导者分工负责的结构。

这段时期他自己格外珍视的经验,是得以同时亲历美国与中国这两个国家的AI生态系统。他判断,美国擅长孕育新的技术构想,而中国则擅长把这些构想转化为产品并迅速推向市场。能够同时踏入这两个拥有不同优势的国家的技术生态系统,他将其视为一段珍贵的经历。这一观察此后具体给他的研究或经营方式带来了怎样的变化,并无资料佐证,因此这里只如实转述他留下的发言。

百度AI团队所做的工作,并不止于撰写论文、发表研究成果。他表示,其团队开发的软件每天被数亿用户实际使用。这个团队的工作大致分为两个方向。其一是支撑百度已经站稳脚跟的核心业务。在他加入之前,百度已经在图像识别、语音识别、自然语言处理和机器翻译领域领先一步,他的团队以此为基础,在搜索、广告、地图服务、外卖、语音搜索、安全、消费金融等公司几乎所有业务领域,推行了数十个项目。他把这些既有业务称为"母船"。这意味着,与其说是在创造全新的技术,不如说更接近于修整一艘已经在航行中的大船的航线、为它提速。

另一件事,则是从零开始创立全新的业务。据他所述,在任职后期的两年里,他每年都会向世界推出一个全新的独立业务单元。他亲自负责自动驾驶汽车团队并使其走上正轨,还以语音识别技术为基础推动智能音箱业务,建立起可以与亚马逊或苹果推出的同类产品相竞争的对话式计算平台。他曾颇为自豪地回忆道,这些团队日后在中国市场也持续取得成果。除此之外,应用于授权人员靠近即可自动开启的闸机的人脸识别技术、医疗领域的对话式人工智能"melody"等新尝试,也都是在这一时期萌芽的。他将自己称为百度AI战略的设计者,这一说法中透着他引领AI在公司内部落地生根这一过程的自豪感。

这三年经历留下的最鲜明的遗产,是他亲身掌握了在大型组织内部将AI融入业务的方法。如果说在谷歌大脑积累的是搭建研究组织的经验,那么在百度,他从头到尾指挥了把研究成果转化为实际营收和产品的整个过程。这段经历后来成为他创立Landing AI、为多家企业撰写AI转型指南的基础。关于在一家企业内部应当如何引入AI、需要以怎样的组织结构来支撑、又该按怎样的顺序扩展业务,这方面的判断力似乎正是源自这段时期的实务经验。

到了第三年,他宣布卸任百度AI团队的领导职务。不同资料对这一时间点的记载相差数日。中文维基百科记载为3月20日,而他本人撰写的辞职文章则于3月22日公开。这数日之差究竟是发表时间与公开时间的差异所致,还是另有行政上的原因,现存资料已无从查证。不过,本章判断以他本人亲笔撰写并向世人公开的文章日期为准,信度最高,因此以他的声明公开的时间点作为辞职的基准。在他之后,百度AI团队由王海峰、百度研究院由林元庆分别接手,延续了组织的运转。

他曾在世界经济论坛的一次采访中亲口谈及,为何要离开这个权力不小、组织也已趋于稳定的位置。他说,自己在审视组织架构图时,渐渐生出一个念头:即便没有自己,团队也能做得很好。已经由一群优秀人才组成的团队,牢牢支撑着既有业务,但他回顾道,自己真正全情投入的时刻,其实是像自动驾驶或智能音箱这样,在一片空白之处从零开始搭建全新业务的时候。在大公司内部延续这样的尝试固然有意义,但他确信,若借助风险投资工作室这种形式,从头创立一项完全崭新的业务,自己的能力将能得到更好的发挥,这成了他辞职最鲜明的动机。至于中国政府的审查问题、与经营层的矛盾、个人家庭事务这类外界饶有兴趣猜测的其他原因,在他留下的任何资料中都无从确认。将这类传闻当作事实来传播,将是一种不够审慎的叙述。

他辞职时所写文章的标题,含有'开启我AI旅程新篇章'之意。在这篇文章中,他再一次表明了自己的核心信念:AI是新的电力。正如近百年来电力改变了众多产业的面貌,AI也将改变医疗、交通、娱乐和制造业等几乎所有主要产业的样貌,让无数人的生活变得更好——这是他的展望。正如过去的工业革命让人从反复的体力劳动中解放出来,他期望如今AI能让人从堵塞道路上握着方向盘这类反复的脑力劳动中解脱出来。

在把谷歌大脑和百度这两家大型企业真正转变为AI企业的过程中,他自我评价说,自己发挥了应有的作用。但他判断,AI所蕴含的可能性,超出了少数几家大型企业的业务范围。离开百度之后,他将这一判断付诸了实际行动。他相继创立了风险投资工作室AI Fund、面向企业的AI公司Landing AI,以及教育机构DeepLearning.AI,致力于将AI技术推广到传统产业的方方面面、让更多人能够学习AI,开启了新的局面。

在百度度过的三年,在他的履历中并不仅仅是壮大一家公司技术实力的时期。这是他亲手掌握把研究转化为实际业务的方法的时期,也是他坚定决心要把这套方法推广到世界上更多企业的时期。离开一个庞大而稳定的位置、重新从零开始创业的选择,似乎是他在学术与产业之间反复往返所做判断的延续。他把在实验室验证过的构想搬上讲台,又把在讲台上遇到的渴求带回产业现场,这样的脚步,伴随着他离开百度时留下的这份宣言,正朝着下一个位置迈进。

离开百度之后,他在森尼韦尔创建硅谷AI实验室的方式,此后成为他所创立的多个组织的原型。他没有选择在美国新设研究据点,而是把已经在北京站稳脚跟的两个研究所收编进来,置于统一的指挥体系之下。与其从零开始搭建一个全新的组织,他更倾向于先把分散的力量汇聚起来、理顺方向。这种方式,与他在谷歌大脑从少数人力起步、逐步壮大组织的经验截然不同。在百度,他的工作从整合已经存在的两个北京研究所与新建的美国实验室开始,这项整合工作本身就是他接下的第一项任务。

这一时期他常常提及的一个词是"规模"。他从谷歌大脑时期起就通过实验证明了扩大神经网络规模能带来性能提升的信念,而在百度,他获得了在搜索这一庞大服务之上重新检验这一信念的机会。把新模型应用到每天有数亿人使用的服务上,这要求承担的是一种与在实验室里改善准确率数值截然不同的责任。服务哪怕只是短暂变慢或出现故障,其冲击都会立刻传导给用户。他所领导的团队之所以能够在搜索、广告、地图、外卖服务等领域同时推进数十个项目,背后很可能是这个组织在反复打磨"把研究成果落地到实际服务"这一流程的过程中积累起来的经验。不过,这一流程具体经历了怎样的验证环节、失败的项目又是如何处理的,现存资料已无从确认。

他回忆称自动驾驶汽车业务和智能音箱业务是每年新创立一个,这让人得以窥见他在组织内部所扮演的角色。管理已有的业务和在一片空白之处创建业务单元,所要求的能力是不同的。既有业务只需沿着既定的目标和指标推进即可,而新业务则必须从头决定要做什么、以怎样的组织形态起步、以什么标准判断成果。他在任职后期每年创立一个这样的新业务的事实表明,他在组织中很可能逐渐把更多时间投入到设计新业务这一角色上,而不是重复性的管理工作。后来他曾表示,比起在大企业内部,以风险投资工作室的形式创立新业务更适合自己,这似乎也与他在这一时期发现的自身倾向不无关系。

在百度内部,他所承担的另一份职责是发掘人才。他在辞职时留下的文章中,一一点出了与自己共事的多位研究者和高管的姓名。以这种方式具体列出同事姓名的辞职信并不常见。这给人留下的印象是,他并未把自己领导组织的这三年只归结为个人的成果,而是试图将其叙述为多人共同创造的结果。其中他尤其评价说,自己发掘的研究者们为夯实组织的知识基础出了力,而这也是他每次创建研究组织时反复采用的方式。无论是创立谷歌大脑,还是后来创立DeepLearning.AI或Landing AI,他都不是选择独自处理一切,而是选择寻找值得信赖的人、分工合作。百度的经验可以说是在更大规模的组织中检验了这种方式。

他将美国与中国的AI生态系统并列比较的观察,并未止步于一种简单的印象式评论,而是在他此后创立的多个组织的方向上留下了痕迹。他判断,美国擅长创造新的技术构想,而中国擅长把这些构想转化为产品并迅速推向市场。反过来看这一观察,也意味着把构想转化为实际产品的速度,是决定技术价值的一个

要素。他离开百度后创立的Landing AI,之所以致力于向企业提供引入AI的具体流程与步骤指导,也有可能是因为他在百度亲身体会到,缩短构想与产品之间的距离有多么艰难。不过,这一因果关系他本人从未直接说明,因此这里只作为一种情理上的关联加以标注。

百度研究院被重组为三个研究所的结构,预先展现出他此后创建组织时反复采用的原则:将分散在不同地区的研究人力统一置于一个指挥体系之下,同时明确划分各研究所负责的领域。可以推测,硅谷AI实验室承担探索新研究方向的角色,而北京的两个研究所则承担把已经积累的图像、语音、自然语言处理技术转化为实际服务的角色。这样的分工虽未见诸文字,但从他所主导的各个项目的性质来看,可以确认研究与应用在地理上也在一定程度上是分开进行的。

三年这一时长本身也值得留意。他领导谷歌大脑的时间不长,担任Coursera联席首席执行官的时间也不长。在百度度过的三年,同样延续了他不长期驻留在某一个组织、而是解决了特定问题之后就转向下一个位置这一模式。这一模式究竟是他有意为之的设计,还是每次面对新机会时顺其自然的结果,无法断言。但从结果来看,他一直是在一个组织中掌握了解决问题的方法之后,便把这套方法带往下一个组织,以此积累自己的履历。百度是这一系列经历中他亲自运营过的规模最大的组织,这段经历也成为他此后能够站在为多家企业AI转型提供建议这一位置上的实际根基。

2014年5月发表任命消息之后,硅谷内外围绕"为何偏偏是百度"这一问题,出现了各种各样的解读。一家在搜索引擎市场上与谷歌相抗衡、虽在美国证券市场上市但总部与营收根基都在中国的公司,聘请一位美国学者出任本公司AI研究总负责人,在此之前并不常见。硅谷媒体将这一人事任命解读为中国企业吸纳美国学术界人才的一个信号,学术界则关注到他以兼职形式保留斯坦福教职、同时统领大企业研究组织的这一步,认为这是一种罕见的兼职结构。不过,这些外部视角对他本人的决定究竟产生了多大分量的影响,现存资料并未显示。

在森尼韦尔开设的硅谷AI实验室,从物理空间上看,就与北京的两个研究所性质不同。在既有业务——搜索与广告——的重心都在北京的情况下,他没有选择在美国土地上新建据点,而是选择把北京已经积累的研究力量纳入自己的指挥之下。这一结构要求他运用此前从未经历过的组织运营方式。在谷歌大脑,他是从少数人力起步、逐步积累成果、壮大组织,而在百度,从一开始摆在他面前的第一项任务,就是整合已经站稳脚跟的两个研究所和新建的美国实验室。这意味着,把地理上分散的组织统合到一个方向上,是他最先着手的工作。

这一时期他需要管理的对象,不仅仅是研究人力。把新技术叠加到搜索、广告、地图、外卖这类已经有数亿人在使用的服务上,要求承担的是一种与在实验室里提升准确率数值截然不同的责任。服务哪怕只是短暂变慢或出现故障,其冲击都会立刻传导给用户。在搜索这一庞大服务之上进行扩大神经网络规模、提升准确率的实验,其分量与谷歌大脑时期在实验室内部扩大规模的工作有所不同。他的团队之所以能够在搜索、广告、地图、外卖、语音搜索、安全、消费金融等广泛领域同时推进数十个项目,背后很可能积累着这个组织在反复打磨"把研究成果落地到实际服务"这一流程过程中所获得的经验。不过,这一流程具体经历了怎样的环节、失败的尝试又是按怎样的标准收尾的,现存资料已无从确认。

支撑既有业务与从零开始创立新业务,所要求的能力彼此不同。既有业务只需沿着已经确定的目标和指标持续改进即可,而新业务则必须每次都重新决定要做什么、以多大的组织规模起步、按什么标准判断成果。他回忆称,在任职后期的两年里,自己每年都向世界推出一个新的业务单元,这让人得以推测,他在这一

时期很可能把越来越多的时间投入到设计新业务这一角色上,而非组织管理者的角色。把自动驾驶汽车团队和智能音箱团队推上正轨的经验、孵化人脸识别技术和医疗对话式人工智能的经验,都集中在这一时期。他亲身确认了即便在大企业这样庞大的组织内部,这样的尝试也是可行的,但与此同时,他似乎也开始对这种方式是否最适合自己产生了疑问。

领导百度的这三年,与他此前走过的职业生涯相比,也是格外引人注目的一段经历。他领导谷歌大脑的时间不长,担任Coursera联席首席执行官的时间也不长。他一直是在一个组织中掌握了解决某一特定问题的方法之后,便把这套方法带到下一个组织中加以应用,以此积累职业履历。百度是这一系列经历中他亲自运营过的规模最大的组织。统领多达1300人的人力、同时兼顾搜索这一核心业务与自动驾驶这一新兴业务的经验,成为他此后能够站在为多家企业AI转型提供建议这一位置上的实际根基。从领导少数研究者的学者,转变为运营数千人规模的组织、同时对营收和产品两端都负责的经营者,这三年的转变,似乎已经化作实务上的分寸感,融入了他日后创立Landing AI时给企业们的建议之中。

他关于同时经历美国与中国这两国AI生态系统的观察,也源自这一时期的经验。他判断,孕育新构想的力量与把这一构想转化为产品的速度,在不同的国家各自成长起来,这也意味着缩短构想与产品之间的距离,是决定技术价值的一个要素。他离开百度后创立的Landing AI之所以致力于向企业提供引入AI的具体流程与步骤指导,背后很可能正是因为他在百度亲身体会到,缩短这一距离有多么棘手。不过,这一因果关系他本人从未直接说明,因此这里只作为一种情理上的关联加以标注。

在宣布辞职时留下的文章中,他一一点出同事姓名的方式,也值得留意。以这种方式具体列出共事者姓名的辞职信,并非常见的形式。这给人留下的印象是,他并未把这三年的成果只归结为个人的功劳,而是试图将其叙述为多人共同创造的结果。这种态度在他此后创立的组织中也反复出现。无论是创立谷歌大脑,还是创立DeepLearning.AI或Landing AI,他都不是选择独自承担所有角色,而是选择寻找值得信赖的同事、分工合作。百度的经验可以说是在规模远为庞大的组织中检验了这种运营方式。

离开百度之时,他似乎已经在一定程度上勾勒出了下一步的方向。他回忆道,自己在审视组织架构图时确信,即便没有自己,团队也能做得很好——这并非只是一位疲惫经营者的吐露,而更像是指向下一阶段的判断。通过在大企业这艘大船之内一个接一个地推出新业务的经历,他确认了自己最全情投入的时刻,正是在一片空白之处从零开始搭建某样东西的时候。这一确认,促使他在离开百度后不久便直接选择创立风险投资工作室这种形式的组织。在借助大企业的资源与规模来试验新业务、与从一开始就作为独立组织创业这两条路之间,他选择了后者,而这一选择的根基,正是他在百度亲身经历过的两件事——管理一个庞大组织,以及在其中培育新兴业务。

百度美国研发中心开业的日期是2014年5月16日。当天发布的新闻稿中具体出现了他将要领导的组织名称。百度把自己整体的人工智能计划命名为"百度大脑",他则出任统领这一计划的负责人。一家运营搜索引擎的公司,给自己的核心技术战略冠以"大脑"之名,这一事实读来意味着,这家公司并未把AI视为附属功能,而是将其当作支撑整个公司的神经系统。这个名称是在他加入之前就已经存在,还是以他的加入为契机而重新确立的,现存资料已无法确认。

从他所领导组织的人员构成来看,可以发现这并非一支只汇集了研究者的团队。他在辞职时留下的文章中写道,自己的团队从经营层到优秀的工程师、科学家、产品经理,由多个层次的人才厚厚地充实起来。运营一个研究所,与运营一个贯穿研究到产品发布全过程的组织,所要求的人员构成本身就不同。只有写

论文的人,无法把技术叠加到每天有数亿人使用的服务上;只有制作产品的人,也无法确定技术的方向。他之所以能够组建起同时承担这两种角色的组织,背后似乎是一种把研究者和产品负责人并列招募、加以配置的眼光。

在辞职信中被提及姓名的人物,不止陆奇、王海峰、林元庆、亚当・科茨。他还一并提到了景鲲、李平、徐伟、朱凯华等研究者,写道他们为夯实百度AI组织的知识基础作出了贡献。在一篇辞职文章中并列写下多达八个人的姓名,这样的例子并不多见。离开组织时留下的最后一篇文章,通常会把大部分篇幅用来梳理自己的成果,而他却把相当一部分篇幅用来一一唱出同事的名字。这种方式给人留下的印象是,他试图把自己描绘成一个曾经走过这个组织的人,而不是一个创建了这个组织的人。

这一时期他所承担的组织规模,是斯坦福课堂或谷歌大脑创立初期都无法比拟的。与和少数研究生一起撰写强化学习论文的时期、或以几名人力起步的谷歌大脑初期不同,在百度,他从一开始就继承了超过千人的人力和已经站稳脚跟的业务结构。执掌这样规模的组织,意味着比起确定研究方向,他要把更多时间用在协调各团队之间的优先级、在已经推进的业务与新启动的业务之间分配资源上。在优先向搜索、广告这类支撑公司营收的业务分配资源的同时,还要为自动驾驶、智能音箱这类尚未产生营收的业务逐步配备人力,这种平衡很可能是他每天都要面对的实务课题。

与他此前走过的道路相比,这段经历标志着一个鲜明的转折点。在卡内基梅隆、麻省理工和伯克利积累的是学术成果,作为斯坦福教授积累的是培养学生的经验。在谷歌大脑,他取得了首次证明大规模神经网络确实能够实际运作这一实验性成就。而百度则在这一切经验之上,又叠加了一项经营课题:要在一家已经成熟的大企业的业务结构之中推进新技术,同时还要培育新业务。从研究者出发的他之所以能够具备运营组织的经营者眼光,这三年的时间可以说发挥了决定性的作用。

纵观他离开百度后创立的多个组织,可以发现他所做的选择,方向上与这一时期所经历的规模问题有所不同。AI Fund从一开始就被设计成扶持多支小型创业团队的结构,而不是一个庞大的组织;Landing AI则以帮助多家企业引入AI的顾问性质组织起步,而非局限于单一公司。他把管理超过千人的经验抛在身后,转而把方向调整为同时照看多个小规模团队。一个曾经运营过庞大组织的人重新选择了小规模单位,这一事实表明,他在百度得出的结论,很可能不是关于如何扩大规模的方法,而是关于规模扩大之后会失去什么的领悟。他回忆称,自己在组织架构图中确认了即便没有自己的位置团队也能顺利运转,这似乎促使他做出判断:比起担任大型组织的管理者,同时照看多个小规模的新起点,更适合自己。

# Coursera与降低在线教育的门槛

2011年,斯坦福大学校园里的一间教室正在悄悄改变世界。那一年,吴恩达在斯坦福大学开设了一门在线机器学习课程,报名人数超过了十万人,这一规模远远超出了在大学校园里能够面对面容纳的学生人数。这门课程成为斯坦福大学率先推出的早期大规模开放在线课程之一,并被记录为世界上规模数一数二的大型在线课程。他开始这项工作时怀抱的目标很明确,那就是让世界上任何地方的任何人都能免费获得优质教育。至于他为何身为一所精英大学的教授却如此致力于打破教育围墙,这一决心背后可能存在的个人纠结或成长经历,目前没有资料可以确认。不过他所设立的目标以及为实现这一目标所选择的方式,通过之后一连串的行动清楚地展现了出来。

吸引十万人报名的在线课程并没有以一学期的实验就此结束。第二年,也就是2012年,吴恩达与达芙妮·科勒共同创办了Coursera。他们的构想是要以大规模的方式在线提供大学水准的课程。两人是经过怎样的过程决定合作创业的,期间是否有过讨论或分歧,现存的资料无从查证。但从结果来看,Coursera的诞生在那一年的教育界引发了不小的震动。《纽约时报》将2012年称为大规模开放在线课程之年,并对这一潮流进行了集中报道。世界顶尖的大学和企业纷纷与Coursera携手,课程、专业证书课程与学位课程也相继叠加了上去。随着时间推移,Coursera确立了世界上规模最大的大规模开放在线课程平台的地位,注册学习者人数超过了1.5亿人。吴恩达一直担任Coursera联合首席执行官到2014年。那一年他转任百度,但他与Coursera的缘分并未就此中断,他仍以董事会主席兼联合创始人的身份继续与公司保持联系。

打造Coursera这个容器固然重要,但与之同样重要的是打磨其中盛放的课程本身。跟随吴恩达名字时间最久的课程是他的机器学习课。这门课程于2012年首次开课,学员评分在五分满分中达到4.9分,累计有超过480万名学习者修读过。随着岁月流逝,人工智能行业的主要编程语言从Octave转向了Python,顺应这一潮流,吴恩达将这门老课程重新编排为由三门课组成的机器学习专项课程。这次改版的核心问题只有一个,那就是如何让不熟悉数学和编程的人也能跟上这些内容。因此新课程选择了先用图形和图表展示概念,再让学员用代码实现算法的顺序。这种方式不是先抛出公式强迫理解,而是先用眼睛把握大致感觉,再动手加以确认。参与这次改版工作的还有埃迪·徐、阿尔蒂·巴古尔、杰夫·拉德维格等人。阿尔蒂·巴古尔是曾在斯坦福和Landing AI与吴恩达共事过的机器学习工程师,埃迪·徐则是在DeepLearning.AI打造过多门教育课程的产品负责人。这些人是通过怎样的契机组成一个团队的,资料中并未留下记录,但这样重新编排的机器学习专项课程,至今仍有超过812608名学习者在Coursera上注册学习。

一门课程留在一个人生命中的痕迹,并非仅凭数字就能说清。曾任数据工程师的奇拉格·戈达瓦特在修读这门课程的过程中成为了机器学习方面的导师,在IEEE发表了研究论文,最终在摩根大通谋得一职。先锋集团澳大利亚分公司的高级交易员尼古拉斯·穆钦古里则利用这门课上学到的内容,着手将投资流程自动化。当时还是学生身份的阿卡什·萨鲁普撰写了一篇关于面部表情识别的论文,并凭借这项成果获得了摩根士丹利的实习机会。其中最引人注目的案例是内克塔里奥斯·卡洛格里迪斯。当时正处于失业状态的他,将修读这门机器学习课程视为人生中最大的转折点,并进一步将深度学习与华尔街数据结合起来,创办了一家初创公司。很难说这些案例能够代表修读过这门课程的数百万人的全体情况,但这几个故事具体地说明了一间在线教室即便没有学位或人脉,也能够开辟出一条新的道路。

机器学习课程站稳脚跟之后,吴恩达选择了同时向两端拓展教育广度的道路。一端是更加深入的专业课程。他与斯坦福计算机科学系讲师基安·卡坦福鲁什、优内斯·本苏达·穆里携手,开设了深度学习专项

课程。这门由五门课程组成的课程面向中级学习者,分量不轻,需要每周投入十个小时、持续三个月的时间。这条路并不轻松,但对那些想要进入人工智能行业的人来说,它成为了一条实实在在的通道,至今已有超过993144名学习者注册了这门课程。

另一端则专门为那些从未写过一行代码的人预留了位置,那就是名为“人人都能学的人工智能”的课程。这门课程面向的是既非程序员也非工程师,而是在企业中工作的管理者、处理政策事务的人,以及那些单纯对人工智能是什么感到好奇的普通大众。这门课程至今已有超过2563631人修读过,由总计七小时时长的四个模块构成。第一周梳理人工智能是什么这一概念,第二周探讨如何在企业内部搭建人工智能项目,第三周审视如何将人工智能引入整个组织,第四周则审视人工智能对社会造成的影响。厘清神经网络或深度学习这类词语实际意味着什么、人工智能眼下能做到什么以及还做不到什么,是这门课程的核心所在。诸如带有偏见的判断、试图欺骗人的对抗性攻击、对就业岗位造成的影响等超越技术本身的问题,也一并被纳入了讨论。这是一项把曾经只属于专家语言范畴的人工智能,拉低到不懂这套语言的人也能加以判断的位置上的工作。

吴恩达和Coursera试图降低的门槛,不只是知识的难度,还有费用这一不容小觑的障碍。前面提到的这些课程的说明文档中,无一例外都设有资助支持这一项目。对于难以承担49美元证书费用或每月30美元订阅费用的学习者,Coursera设有资助支持制度。只要提交申请并获得批准,便能免费获得与付费学习者完全相同的课程和证书。如果没有这样的制度,再精心打造的课程,能够叩开其大门的人从一开始也只会局限于那些手头宽裕的人。

关于这种降低经济门槛的做法实际改变了一个人生活的案例,吴恩达曾亲自讲述过一段经历。他与谷歌合作开设TensorFlow专业证书课程时,谷歌为这个项目一年投入的资金大约是十万美元。相较于一家巨头企业的财务规模而言,这并不是一笔很大的金额,但这笔小小的投入所创造出的信任与善意,却远远超出了这个数额本身。吴恩达讲述了他在阿姆斯特丹一场活动上遇到一位叙利亚青年的故事。在内战动荡、整个国家风雨飘摇的境况下,这名青年成为叙利亚国内最早获得TensorFlow证书的人之一。这张证书成为他摆脱贫困的踏板,他随后前往德国,在一家大企业谋得了职位。这个故事令人感动,但终究只是一个人的案例。不能仅凭这一个案例就断定世界各地都发生着同样的事情。不过,这个故事足以作为具体的证据,说明一张在线证书在某些国家确实可能改变一个人生活的轨迹。

始于2011年斯坦福大学一间教室的这股潮流,历经Coursera和DeepLearning.AI,至今已向超过800万人传递了人工智能教育。没有依据能够证明这一数字均衡地惠及了世界各个地区,关于究竟触及了哪些国家、哪些阶层、程度如何,现有资料并不充分。尽管如此,这股潮流所展现出的东西是明确的,那就是一门始于单间教室的课程,跨越了大学的围墙,跨越了语言与收入的差距,最终延伸到了每一位学员具体的人生之中。吴恩达在斯坦福大学创办的这一门在线课程,延伸成了Coursera这家公司,又延伸成了这家公司内部的诸多课程,再延伸成了修读这些课程的数百万人各自的人生故事。下一章将要谈到的DeepLearning.AI,正是这股潮流跨越Coursera延伸到另一个组织的节点。

若细看Coursera这个容器内部实际课程是如何运作的,便能更清楚地看出降低门槛这句话具体意味着什么。一周分量的课程被拆分成多个十分钟左右的短视频陆续上线,每段视频结束后都附有一道简短的确认题。中途跟不上进度的人可以当场倒回视频重新观看,再做题确认自己是否真正理解了内容。在大学教室里举手提问的机会往往只有一次,但在线上,学员可以反复回看,直到真正理解为止。在涉及编程的机器学习或深度学习课程中,每周结束时都会布置实际编写并提交程序的作业,这些作业会由系统自动评分,结

果立即返回。无论是坐在斯坦福校园里听课的学生,还是在地球另一端仅凭一台笔记本电脑连接进来的人,得到的评分与反馈都没有区别。这套体系用机器即时给出答案取代了让助教一个人手工翻阅答卷、按顺序排队等候的方式。此外,每门课程都单独设有讨论区,让在同一个问题上卡住的学员之间互相提问、共同寻找答案。既然一位教授不可能一一照顾数十万名学生,于是便搭建了让学员相互教学的场所,来填补这一空缺。

修读Coursera课程有两种方式。一种是不付费,只浏览讲课视频和资料的旁听方式,另一种则是付费提交作业、获取正式证书的方式。仅凭旁听也足以获取知识,但不会留下可以写进履历或求职材料的凭证。反过来,若想获得证书,则需要根据课程支付一次性49美元的费用,或每月30美元的订阅费。前面提到的资助支持制度正是用来消除这道界限的装置。提交申请并获批的学习者,能够获得与付费者完全相同的课程、作业以及同样的证书。这相当于把获取知识的门和证明这份知识的门分开设置,并让后一道门即便对经济困难的人也保持敞开。如果没有这种双重结构,无论课程多么出色,需要证书的求职者或即将换工作的上班族,最终也只会局限于那些手头有钱的人。

机器学习专项课程和深度学习专项课程之所以被拆分为多门课程而非单独一门,也是有原因的。机器学习专项课程由三门课组成,深度学习专项课程由五门课组成,这是为了让学员一步步拾级而上,而不是把庞大的内容一次性抛给他们。前一门课中学到的概念成为下一门课的前提,最后再以一个整合所学内容的项目收尾。即便是不熟悉数学和编程的人,只要修完第一门课,便已经具备了修读第二门课的准备,各阶段的难度就这样被紧密地衔接起来。中途失去兴趣或因其他事情腾不出时间的学习者,即便只修完一门课程,也能在其中获得一份完整的证书;而若条件允许,也可以继续修读下一门课,最终完成整个专项课程。这种设计没有从一开始就强行把五门课捆绑成一整块,而是让学员能够按照各自的情况分步前行。

这样打造出来的课程若要在实际就业市场中发挥作用,还需要有认可这份证书的大学和企业与之相配合。如前所述,Coursera与世界顶尖的大学和企业携手,叠加上了课程、专业证书以及学位课程。谷歌与吴恩达合作打造的TensorFlow专业证书,同样是在这一结构之下诞生的成果。同时挂着大学名号和企业名号的证书,与修完一门无名在线课程所具有的分量截然不同。这为招聘负责人提供了辨别的依据,判断求职者贴出的是在陌生网站上随意制作的材料,还是完成了经过验证的大学和企业共同设计、评分的课程。前面提到的奇拉格・戈达瓦特、尼古拉斯・穆钦古里、阿卡什・萨鲁普、内克塔里奥斯・卡洛格里迪斯之所以能够转化为实际的就业、实习和创业机会,背后正是这套认可体系在起作用。若课程本身再出色,却没有可以信赖的依据来证实这一成果,那也不过是履历上的一行字而已。

这一切潮流所具有的分量,若把规模拿来比较,便会更加鲜明。在斯坦福等名校里,一学期修读一门课程的学生,多则也不过数百人的规模,教室大小以及一位教授所能承担的批改量,决定了这个上限。2011年涌向吴恩达在线机器学习课程的十万人,是斯坦福大学一百多年来所维系的教室容量,单凭一门课程就轻松超越的数字。而且这股潮流并未止于一次性的事件,而是转移到了Coursera这家公司,确立为至今已向超过800万人传递人工智能教育的体系。一所大学历经数百年培养出的毕业生人数,被一个在线平台在短短十几年间超越了。这一对比并非要贬低大学教育的价值,而只是具体的数字,展示了当剥离掉教室这一物理界限时,教育能够以何等不同的规模扩散开来。

从机器学习、深度学习到人人都能学的人工智能,吴恩达打造的这些课程,归根结底都回到了同一个问题上,那就是横亘在求学者与知识之间的障碍究竟能降到多低。为了没有数学背景的人,他先用图形和图表来展示;为了经济困难的人,他设立了资助支持;为了在就业市场上获得信任,他把大学和企业的名号一并

挂了上去。打磨课程内容的工作,与打磨围绕课程的制度的工作是并行推进的。在Coursera积累下来的这份经验,原封不动地延续到了吴恩达之后创办的另一个教育组织之中。在大学这把大伞下诞生的课程,如今脱离这把伞、以独立的名号立足,正是这股潮流的下一个阶段。

在Coursera这个名字出现在世人面前之前,吴恩达教授的那门在线课程在斯坦福内部被称为大规模开放在线课程这一新名称。大规模一词指的是注册人数的规模,开放一词则意味着修读这门课程不需要入学考试,也没有名额限制或学费。大学几个世纪以来所坚守的大门前,一直摆着入学审核和学费这两道关卡。吴恩达在2011年创办的这门课程,正是一次性撤掉了这两道关卡的场所。只要有网络连接和一台笔记本电脑,任何人都能听到斯坦福教授的授课,正是这一个条件造就了十万这个数字。《纽约时报》将2012年称为大规模开放在线课程之年,其背后并不只有斯坦福一家。同一时期,其他多所大学和教育机构也相继推出了同样方式的大规模在线课程,媒体将这整股潮流合并当作一个事件进行了报道。吴恩达和达芙妮・科勒创办的Coursera,只是这股潮流中的诸多尝试之一,但随着时间推移,它成长为世界上数一数二规模的案例留存了下来。

创办Coursera的决定,与持续开设一门在线课程是不同性质的决定。学校即便按照每学期重新开课的方式,也能每次招收十万名学生。然而吴恩达和达芙妮・科勒并没有把这门课程留作单一大学的一次性活动,而是新打造了一个能让多所大学和多个机构共同叠加课程的公司这一容器。凭借这一决定,Coursera不再只是一个反复播放斯坦福单一课程的地方,而成为了汇聚世界多所大学课程的场所。达芙妮・科勒同样是斯坦福计算机科学系的教授,一直在讲台上教授学生,两人共同创办的这家公司,起到了连接大学这一古老制度与在线这一新通道的桥梁作用。单靠一所学校的名号无法承载的规模,以多所学校共同参与的平台这种形式承载了下来。

Coursera不止步于大学,还与企业携手,这一事实展现出在线教育正从学术领域跨入产业领域的节点。传统的大学课程以学位为目标而设计,认可这份学位的主体是其他大学和学术界。Coursera上叠加的多项证书课程脱离了这一框架。像谷歌这样的企业直接参与设计证书这一事实,意味着评判这份证书的标准不再是学术界的认可,而是就业市场的需要。在大学历经长时间构筑起来的学位体系旁边,又并列出现了另一套能在短期内证明企业所需实务技能的体系。这两套体系并非相互排挤的关系,而是并存共立的关系。想要获得学位的人依然去大学,而想要迅速掌握某项特定技能并留下证明的人,则会去寻找Coursera的证书课程。

2014年吴恩达转任百度、卸任Coursera联合首席执行官一职,并不意味着他与这家公司的关系就此终结。他仍留任董事会主席兼联合创始人一职。虽然放下了公司日常运营的工作,但他在公司创立之初所确立的方向上,依然挂着自己的名字。他把经营交给了他人,自己则留在了更根本的位置上,也就是决定教什么内容、以何种方式来教的位置上。事实上,他之后打造的机器学习专项课程和深度学习专项课程,都是以课程设计者而非经营者的名义问世的。他把公司日常运营与课程内容这两件事分开处理,前者交给了其他管理层,后者则由自己继续把持着。

1.5亿这一注册学习者规模,以及800万这一人工智能教育修读者规模,都超出了任何一个国家的大学体系所能承载的数字。这一规模所意味的是,经由Coursera的人并不局限于斯坦福或美国国内。只要有网络连接,任何地方都能听到同样的课程,这门教育从一开始就等同于是没有把国界放在设计考量之中的。诞生于硅谷、在硅谷成长起来的技术,由此打开了一条向硅谷之外扩散的通道。处理尖端技术的学问,往往呈现出只在少数聚集了研究经费、实验设备以及能教授该领域的教授的城市里才能发展起来的倾向。

吴恩达打造的这门在线课程,正是一条用一门课程绕开这种地理限制的道路。没有优质大学的地区、没有教授人工智能的教授的国家,如今也能凭借一台笔记本电脑听到斯坦福水准的课程。

随着这种规模的教育扩散开来,自然而然随之而来的问题是,在线课程是否会取代传统的大学教育。Coursera实际走过的道路,给出了对这一问题的一种回答。Coursera并没有试图取代大学,而是选择了与世界顶尖大学并肩携手的道路。课程虽然是在Coursera这一平台上开设的,但制作这些课程的地方依然是斯坦福等大学。大学提供课程内容和名号,Coursera提供让这些课程能在世界任何地方被看到的通道,这样一种结构确立了下来。在这一结构中,大学并未消失,反而是大学这一名号的分量借助在线这一新通道,触及到了更广阔的地方。原本只在校园内部通行的大学权威,如今延伸到了校园之外的数百万人身上。

设有免费开放课程供人修读的结构,与需要付费才能获得正式证书的结构并存这一事实,揭示出接受教育这一行为与证明这份教育之间是两件不同的事情。获取知识本身不需要花钱,但向他人证明自己获得了这份知识,依然以某种形式被标上了价格。这种双重结构与其说是Coursera独有的发明,不如说更接近于在线教育大规模扩散过程中自然而然确立下来的一种方式。它让任何人都能够接触到知识本身,同时也让想要向世界证明这份知识的人,需要经过最基本的一道程序。这道程序虽然不是完全免费的,但通过资助支持制度,同样的证书也向经济困难的人敞开了大门,由此知识和证明这两样东西,都留有了不花钱也能获得的途径。

回顾这股始于2011年斯坦福一间教室的潮流,吴恩达通过Coursera所成就的事情,并不仅止于把一门课程做大。它留存下来的,是一个展现大学这一古老制度遇上在线这一新通道后如何能够扩展自身影响力,以及在这一过程中企业与学术界如何能够并肩合作的案例。即便在他从经营岗位上退下、只留任课程设计者的身份之后,Coursera仍持续成长,而他之后创办的组织,也原封不动地延续了这一时期打磨出来的方式。从一间教室出发的实验,发展成一个由世界多所大学和企业共同参与的体系,这一过程的核心始终萦绕着同一个问题,那就是知识究竟能被送到多远的地方。

Coursera这家公司与大学携手叠加的课程,并非只有单门课程或几门课打包而成的专项课程。随着时间推移,这份名单上还加入了正式的硕士学位课程。与在线上几周内就能结束的证书不同,学位课程是把大学发放的毕业证书本身以在线路径开放出来的场所。原本需要在校园里缴纳学费才能获得的毕业证书,如今通过网络接入的学习者也能获得同样的名号。在这一结构下,Coursera把证书课程和学位课程并列摆放,学习者可以根据自己的目标,选择几周就能完成的证书,也可以选择要花几年时间的学位。想要靠一张短期证明来准备换工作的人,与想要凭一份正式学位重新累积职业生涯的人,在同一个平台上各自找到了不同的门径。

在一家公司内部叠加起如此多层次的课程,同时也是一项在免费分享知识的承诺与公司必须持续运营这一现实之间寻求平衡的工作。既要把讲课视频和资料免费开放,又要让公司不至于关门,就必须有某处资金流入。Coursera选择的答案,是不对知识本身收费,而是对证明这份知识的程序收费。它向想要拿到证书、资格证或学位这类文件的人收取费用,而对不需要这份文件的人,则原样开放讲课内容。资助支持制度则是一道例外装置,面向那些无力承担这笔费用的人,免费发放同样的文件。公司的经营与免费教育这一目标,就这样各自在不同的位置上分担各自的角色,共同运转了下去。

“人人都能学的人工智能”课程所瞄准的对象并非程序员这一事实,恰恰体现了这门课程所处理问题的性质。决定要不要把人工智能引入企业的人,大多不是编写代码的工程师,而是掌握预算的管理者。该在

哪个部门优先引入人工智能、交给外部企业承接还是内部组建团队,这类判断往往掌握在不懂技术的人手中。这正是这门课程第二周探讨如何在企业内部搭建人工智能项目、第三周探讨如何将人工智能引入整个组织的原因所在。这门课程没有教代码,而是教会不懂代码的人也能提出判断人工智能项目成败的问题。数据是否充足、问题定义是否明确、期望的准确率是否切合实际,这类问题即便不是工程师也能学会,而学会这些问题的管理者,便能自行判断自己公司里哪些项目该推进、哪些该放弃。

把机器学习专项课程和深度学习专项课程放在一起来看,还能发现这两门课程所瞄准的学习者之间存在层级。机器学习专项课程是即便不熟悉数学和编程的人也能开始的入口,深度学习专项课程则是通过这道入口的人接下来要迈向的场所。由于顺序编排得当,修完一门课程后能自然而然衔接到下一门课程,学习者便能在Coursera内部自行估量自己目前站在哪个位置、接下来该学什么。如果是在大学,原本会用年级和先修课程这类制度来编排这个顺序,而Coursera即便没有这样的制度,也仅凭课程的排列顺序,发挥了相同的作用。初学者可以毫不畏惧地进入,而积累了实力的学习者也有可以继续迈向更难课程的去处。

在这一结构中值得关注的一点是,与吴恩达一同设计课程的人们并没有只停留在Coursera一家。阿尔蒂・巴古尔也曾在Landing AI与他共事,埃迪・徐则在DeepLearning.AI打造了多门教育课程。始于Coursera的这种协作,原封不动地转移到了他之后创办的其他组织之中。这既说明他离开Coursera之后依然没有放下教育这份工作,也说明与他信赖的人们一同打造课程的这种方式,即便在辗转不同组织之后也没有发生太大的变化。共同编排机器学习专项课程的那双手,延续成了编排他之后创办的组织所开设课程的那双手。

Coursera之所以能够触及世界各地的学习者,其基础并不只有网络这一个条件。能否看懂并理解课程的语言障碍、是否有余钱支付证书费用、以及该国是否存在能够认可在线所学内容的就业市场,这些因素都交织在一起。在前面提到的叙利亚青年的案例中,他所获得的不只是课程内容。正因为有谷歌这一名号挂在证书上,他才没有仅凭出身于经历过内战的国家这一点就被材料筛掉,进而在德国一家企业获得了职位。知识借助网络跨越了国界,但证明这份知识的文件所获得的信任,只有借助谷歌或斯坦福这样的名号才能跨越国界。Coursera之所以在每门课程上都挂上大学和企业的名号,原因就在于此。

这份名号的价值,并非单凭Coursera一己之力打造出来的。斯坦福这个名号先前已经积累了百余年的信任,谷歌这个名号也在硅谷积累了相应的业绩。Coursera所承担的角色,是把这份信任移植到在线教室这一新容器之中。它把大学和企业历经长年累月建立起来的信用,分享给了经济困难的学习者和远程学习者。吴恩达在Coursera所成就的成果,并不是创造出了新的知识,而是把原本就已存在的知识以及支撑这份知识的信任,移交给了那些原本难以获得它们的人。

# DeepLearning.AI:面向所有人的AI课程

吴恩达离开Coursera之后,在线教育并没有从他的工作中消失,反而变得更加广阔。2017年,他创立了一家名为DeepLearning.AI的教育科技公司。这家公司提出的使命可以简单概括为:通过顶尖水平的教育、实务培训,以及能让学习者彼此连接的社区,培养全世界的人才,使他们能够共同塑造由AI引领的未来。这话听起来很宏大,但公司实际做的事情却很具体,就是不断制作课程,尽可能降低价格让更多人能够听到,并持续更新内容。

DeepLearning.AI的出现有其前史。2011年,吴恩达在斯坦福大学在线公开了一门机器学习课程,报名人数超过十万。这一刻证明了走出大学教室的课程也能聚集如此多的人。第二年,也就是2012年,他与达芙妮・科勒一起创立了Coursera,把这场实验扩大到公司规模。DeepLearning.AI是下一步。如果说Coursera更像是汇集众多大学和机构课程的市场,那么DeepLearning.AI则是以吴恩达本人及其团队亲自制作的课程为核心。从只需一小时的实操型课程,到需要数月才能完成的专业证书课程,如今以这家公司名义推出的项目已超过150个。经由这些课程的学习者超过700万人。这一数字是DeepLearning.AI自身公布的,需要说明的是,它可能因统计时点和方式而有所不同。

这家教育公司从一开始并没有只面向开发者,反而先把一门几乎相反的课程作为代表作推出。AI for Everyone,译成中文就是《面向所有人的AI》。这门课程的对象不是工程师,而是从未写过一行代码的人、需要在公司里与AI团队协作的非技术岗位员工、负责规划产品的管理者、市场营销人员,以及决定公司方向的经营层——那些需要理解AI但不必亲自动手制作AI的人。整门课程总时长约六小时五十分钟,由三十五个视频讲座和四份评分作业构成。不需要懂数学,也不需要懂编程就能学习,这正是这门课程存在的理由。

这门课程分为四周。第一周从追问AI是什么开始,梳理机器学习、深度学习、数据科学、神经网络等词语各自的含义,更重要的是区分出如今的AI实际能做的事和做不到的事。人们对AI的许多误解,正是源于分不清这两者。缩小AI像人一样思考的想象与其实际上是从海量数据中发现模式的技术之间的距离,是第一周的目标。

第二周讲的是AI项目实际上是如何运转的,展示从想法出发,到收集数据、训练模型,再到把它部署到实际服务中的整个流程。这一部分对非技术岗位的学习者尤其重要,因为如果在公司里负责发起或批准AI项目的人完全不了解这一过程,就很难判断需要多长时间、需要什么条件。第三周讲的是如何让AI在组织内部落地,说明如何与AI团队协作、如何找出公司内部有机会应用AI的问题,以及如何制定不止于一次性项目的可持续AI战略。第四周把目光从技术转向社会,讨论AI带来的歧视和偏见问题、试图欺骗系统的对抗性攻击、发展中国家所处的AI境况,以及AI对就业的影响。一门讲授技术的课程把最后一周整整留给伦理与社会问题,这很能说明这门课程的性质。

听过这门课程的人也留下了反馈。一位从事软件质量管理工作的学员评价说,这门课程广泛涵盖了AI的核心主题,同时从项目选择阶段到执行阶段都提供了清晰的指引。另一位学员表示,借助真实案例和案例研究,自己得以理解AI在多个行业中实际产生的影响。这些评价共同指向一点:没有用复杂的语言去解释复杂的技术。

如果说AI for Everyone是降低门槛的课程,那么Deep Learning Specialization就是为跨过这道门槛的人准备的课程。这个被称为深度学习专项课程的项目由五门课程组成,属于中级系列,面向那些真正想在AI产业最前线开发技术的人。这个课程起到基础课的作用,帮助学习者理解深度学习能做什么、有哪些局限,以及这些成果在实际产业中如何被使用。课程以Python和TensorFlow为基础展开,涉及医学影像判读、自动驾驶、手语识别、音乐生成、自然语言处理等实际应用案例。它不停留在教室里的理论,而是不断展示这项技术究竟被用在了哪些地方。

学习者在这门课程中会亲手搭建并训练几种关键的神经网络结构:擅长处理图像的卷积神经网络、处理有序数据的循环神经网络、能在长序列中不遗忘并保持信息的长短期记忆网络,以及成为近来大规模语言模型基础的Transformer。除此之外,还会学习一些提升模型性能的实务技巧,比如防止过拟合的dropout、稳定训练的批量归一化,以及合理设置初始权重的初始化方法。这些技巧并非只出现在论文中的概念,而是每次训练模型时都会遇到、需要解决的实际问题的工具。深度学习专项课程所涵盖的范围,可以看作是尽力缩小理论与实务之间差距的一种尝试。

随着时间推移,AI技术的重心也发生了转移,从识别和分类图像的时代,进入了像与人对话一样写文章、生成答案的生成式AI时代。DeepLearning.AI顺应这一变化推出了新课程,Generative AI for Everyone,即《面向所有人的生成式AI》。这门课程同样是不要求编程技能的入门课程,面向商业领袖、专业人士以及普通大众。全长五个多小时的课程由吴恩达亲自讲授,梳理生成式AI的运作原理、能做到什么、又从哪里开始做不到。

这门课程也分为三周。第一周夯实生成式AI的基础,展示写作、阅读、对话等日常用途。第二周不止步于提问,而是学习将检索结果结合进模型以提高答案准确性的方法、针对特定目的对模型进行再训练的微调等技巧,并考察一个AI项目从开始到结束会经历怎样的生命周期。第三周放宽视野,探讨生成式AI对商业和社会产生的影响,包括如何打造负责任的AI,以及通用人工智能这一遥远目标与我们当下所处位置之间的关系等问题。学完这门课程的学习者,能够把生成式AI引入自己的日常工作,制定提升生产力的策略。

同一时期推出的另一门课程则正面瞄准了开发者,即ChatGPT Prompt Engineering for Developers,《面向开发者的ChatGPT提示工程》。这门课程与OpenAI合作制作,是一门在一小时四十分钟这样短的时间内传达核心内容的入门短期课程,目标是帮助开发者利用大规模语言模型更快地构建新的应用程序。课程涉及的核心技术包括把文本简要整理的摘要、从上下文中读出含义的推理、把一种格式的文本转换成另一种格式的转换,以及把简短文字丰富扩充的扩展。这四项是任何处理提示词的人都会遇到的基本工作,这门课程教的正是如何组织出能有效完成这些工作的句子。

DeepLearning.AI的课程列表并未止步于此。由于AI技术以数月为单位不断变化,这家公司也持续与合作伙伴一起制作短期课程,好让从业者能够快速掌握新工具。至于与各家合作伙伴结缘的契机、公司内部的决策过程,现有资料无法确认,但从结果呈现的课程列表可以清楚看出这家公司正朝着哪个方向前进。

首先是降低编程本身门槛的课程。AI Python for Beginners时长十一个半小时,专为完全没有编程经验的人设计,教他们借助AI的帮助编写、测试并修正Python代码。学完这门课程,就能达到独立制作简单AI应用程序的水平。在学习编程的过程中把AI当作助手来使用,这一构想再次降低了编程教育的门槛。类似取向的课程还有Spec-Driven Development with Coding Agents,这门课程与JetBrains合作制作,讲授

如何写出让编程智能体能够据以维护软件的清晰规格说明书。这门课程说明,与其亲自动手写代码,不如向AI智能体下达精确指令,而这种能力正在成为一项新技术。

接下来是更深入探讨最新语言模型的课程。Reasoning with o1与OpenAI合作制作,教授如何针对需要复杂推理的任务对OpenAI的o1模型进行提示和使用,内容涵盖如何制定计划、如何编写代码、如何对图像进行推理,以及针对提示词本身进行设计的所谓元提示技巧。Fast and Efficient LLM Inference with vLLM是与Red Hat合作的中级课程,讲授如何优化开源语言模型,并利用名为vLLM的工具实际部署并衡量性能。教育范围已经从教人如何制作模型,扩展到教人如何让模型在实际服务中快速且经济地运行。

最近的趋势是围绕能动式AI,也就是能够自主判断和行动的AI系统展开的课程。Multi AI Agent Systems with crewAI时长三个多小时,讲授如何用自然语言设计并指挥由多个AI智能体组成的团队,这是一种由分工不同的多个智能体协作,来自动化处理单一语言模型难以独自完成的复杂工作流程的方式。Building Multimodal Data Pipelines由Snowflake合作制作,讲授如何把图像、音频、视频等数据转换成语言模型可以处理的文本形式,并在此基础上构建能同时处理多种形式数据的应用程序。Agent Memory: Building Memory-Aware Agents是与Oracle合作的课程,教授如何赋予智能体一套记忆系统,使其能够在多次对话会话之间存储知识,并在需要时取出并加以完善。制作出随时间推移能自我学习、自我改进的智能体,如今已经进入了实务教育的领域。Build and Train an LLM with JAX由Google合作制作,是一门利用Google的开源库JAX,从头到尾亲自训练一个拥有两千万参数的语言模型的课程。它不止步于学习如何使用庞大的商用模型,而是提供了亲手制作模型本身的经验,哪怕规模较小,这一点与其他课程有所不同。

这样罗列下来,DeepLearning.AI的课程列表看起来就像一道阶梯。完全不懂AI的人从AI for Everyone迈出第一步,想学编程就转向AI Python for Beginners,想扎实打好深度学习基础就学Deep Learning Specialization,想掌握生成式AI时代的工具就经过提示工程系列课程,最终抵达能够设计智能体系统的水平。这道阶梯的每一级由不同的合作伙伴、不同的技术、不同的难度构成,但目标只有一个:让AI这项技术不再是只有少数研究者才能掌握的东西,而成为任何愿意投入关注和时间的人都能拿到手的工具。

吴恩达在许多场合都把AI比作电力。一百年前电力被发明时,这项技术起初只是流经少数工厂和实验室的特殊力量,但没过多久,电力就变成了流入每个家庭、每座工厂的理所当然的基础设施。DeepLearning.AI制作的这些课程,可以看作是在教育这个领域实现这一比喻的尝试。只要AI还停留在只有少数专家才能掌握的技术,它的力量就无法扩散到整个社会。每一门课程本身或许算不上什么了不起的大事,但从入门到高级、从非开发者到追赶最新技术的开发者都囊括在内的这整份列表,读起来就像是一套试图把AI这股力量输送到社会各个角落的体系。

当然,现在说这套教育体系已经完成还为时过早。一门课程刚推出,几个月后新的模型和新的工具就会出现,DeepLearning.AI又得追在后面制作新课程。面向从业者的最新课程列表不断增加,原因也在于此。追赶技术的速度不是一次成就就能完结的事,而是每次都要重新开始的事。尽管如此,这家公司迄今为止坚持的原则始终如一:在解释困难概念时不再堆砌更困难的语言,并且为不懂编程的人、刚开始学编程的人,以及已经在业内实际操作模型的人,都各自留出一扇能从自己所在位置开始的门。

DeepLearning.AI能够建立起如此密集的课程列表,其制作课程的方式本身也值得关注。吴恩达每次筹备新课程时都会先确定合作企业,OpenAI、Google、Snowflake、Oracle、Red Hat、JetBrains等公司都名列其中。这些合作究竟经过怎样的具体流程促成,与各家公司签约之前又经历过怎样的协商,现有资料无法确认。不过从结果呈现的课程来看,有一种方式清晰可见:制作新工具的公司与讲授该工具的教育机构携手合作。优化vLLM的课程与经手vLLM的Red Hat合作制作,用JAX训练语言模型的课程则与制作JAX的Google合作制作。如果制作工具的地方和讲解工具的地方不同,两者之间就可能出现信息滞后或失真的空间,而这家公司选择的方式正是缩小这道缝隙。

得益于这种合作结构,每门课程的寿命或许不长,但课程所涵盖的技术却接近实际业内使用的最新版本,这意味着学习者能把课上学到的工具直接搬到自己的工作中使用。这一点与大学教室里进行的正规课程体系不同。大学的教学大纲以一学期、一学年为单位编排,不易更改,而DeepLearning.AI的短期课程则以几周到几个月的间隔不断推出又下架。有些课程推出没多久,所讲授的模型就已经过时。即便如此,这家公司坚持的做法不是撤下旧课程,而是不断叠加新课程,这种结构为想听旧课程和想听最新课程的学习者都留下了选择。

课程的长度也是一个引人注目的特点。有像AI for Everyone这样六个多小时的课程,也有像Multi AI Agent Systems with crewAI这样三小时的课程,还有像ChatGPT Prompt Engineering for Developers那样仅一小时四十分钟的课程。课程被切分成学习者一两天内就能完成的分量。对于要一边上班一边挤时间学习的人来说,这样的短课程比动辄几个月的课程更接近能够真正学完的形式。这种旨在提高完成率的设计,与直接照搬大学课堂的早期在线课程走的是不同路线。把课程切短,就意味着要相应地更频繁、更大量地推出课程,而这份负担完全落在了DeepLearning.AI这个组织身上。

把这些课程所涉及的技术脉络按顺序排开,也能同时读出过去几年AI产业移动的方向。在图像分类和语音识别等深度学习技术占据一极的时期,教授卷积神经网络和循环神经网络的课程是中心。此后随着能生成文章的大规模语言模型登场,教授如何撰写提示词的课程增多。而最近,讲授单一语言模型不再独自工作、而是由多个智能体分工协作的方式,以及智能体记住过去的对话并据此改进判断的方式的课程不断增加。课程列表的变化本身,已经成为记录AI技术编年史的文献。至于吴恩达是提前预见到这股潮流才规划课程顺序,还是先跟随市场上出现的工具再制作课程,已无从确认,但留存下来的列表清楚地表明:每当技术发生变化,教育都没有落后,而是紧紧跟了上去。

DeepLearning.AI制作的课程之间还有另一个共同点:无论哪门课程都不止步于理论讲解,一定会以实操或编程作业收尾。就连像AI for Everyone这样不要求编程的课程,也通过四份评分作业让学习者自行梳理所学的概念。Deep Learning Specialization在理论讲解之后,必定跟着一项亲手实现神经网络的作业。这种结构看起来是为了弥合听课与真正会用之间的差距而设置的装置,整个课程设计的前提都建立在一点上:懂得某个概念的人,和能用这个概念做出成果的人是不同的。

仅凭课程列表和实操结构,还不足以说明DeepLearning.AI所做的全部工作。除了教室形式的教育之外,这家公司还每周发行一份名为The Batch的新闻通讯。这份通讯与有固定教学大纲和评分作业的课程性质不同,内容以汇总当周AI业界发生的消息、梳理解说新发表的论文或模型为核心。如果说课程一旦制作出来,几个月甚至几年都保持同样的内容,那么这份通讯讲的就是当周、当月发生的事。课程难以填补的时间差,由这份通讯来弥补。从学习者的角度看,由此形成了一种结构:通过正规课程掌握基础和技术,再通

过通讯追上这项技术此刻已经走到哪一步。两者以不同的速度运转,但最终指向同一个目的,那就是让学过的人不要停止学习。

这份通讯所涉及文章的性质也有值得关注之处。它不止于罗列已发布的事实,还附带解说该消息在实务中具有什么意义。一篇发布新模型时只传达其性能数据的报道,与一篇说明该模型将如何改变开发者实际工作方式的文章是不同的。The Batch更接近后者,它读完论文和发布资料之后,梳理并传达那些想把内容运用到实务中的人需要知道什么。这一点与DeepLearning.AI在制作课程时坚持的态度是一脉相承的,即不是原样传递信息,而是先设身处地地考虑接收信息的人需要什么。

DeepLearning.AI最初提出的使命中,除了教育和实务培训之外,还有一项,那就是打造一个能让学习者彼此连接的社区。这一部分只看课程说明并不容易察觉,但它从一开始就写在公司自己公布的目标语句里。在线课程本来更接近独自听、独自结束的形式,看完视频、提交作业往往就算完成。DeepLearning.AI想在这种结构里加入让人与人相遇的空间,一个让上课的人彼此提问、一起解决卡住的地方,进而彼此在职业发展上给予帮助的空间。没有这样的空间,课程就只是传递知识的通道;有了这样的空间,课程就同时也成了连接人与人的通道。这家公司把这个社区置于使命语句的前面,而不是课程本身,这一事实说明这个组织把人与人之间的关系,看得和知识传递同等重要。

学完课程的学习者最终留下什么,这一点也值得细看。深度学习专项课程以及多个专业证书课程,结业时会一并颁发结业证书。这份证书不像学位那样由国家或大学背书,而是作为在简历或求职活动中展示自己掌握了何种技术、达到何种水平的依据来使用。如果负责招聘的人在应聘者简历上看到DeepLearning.AI的课程名称时,能推测出这个名字意味着什么,这份证书就发挥了它应有的作用;反过来,如果没有人认得这个名字,这份证书就只是一张纸。DeepLearning.AI精心设计每一门课程的内容和难度,原因之一正是这份证书终究要成为在业内获得信任的标志。制作课程这件事,与守住这门课程名号的价值这件事,并不是彼此分开的。

把这一切活动合在一起看,DeepLearning.AI这个组织的位置就变得更加清晰。这家公司与其说是卖课程的地方,不如说更接近是通过课程、新闻通讯、社区、结业证书这几项装置,让人们持续留在AI这项技术身边的地方。一个听完一门课就结束的人,和一个听完课之后订阅通讯、还在社区里持续与其他学习者交流的人,即便经过同样长的时间,最终也会站在不同的位置。吴恩达在设计这几项装置时,是按照怎样的顺序先决定了什么,现有资料难以确认,但从目前留下的成果来看,能看出他在课程这一种形式无法填满的每个位置上,都逐一添加了不同形式的装置的痕迹。让学习的人不停止学习,DeepLearning.AI通过多种形式的产出反复尝试做的事,最终都收拢到这一件事上。

把课程列表摆开来看,还有另一处标记引人注目,那就是初级和中级这样的难度标注。AI for Everyone标注为初级,Generative AI for Everyone也是初级,ChatGPT Prompt Engineering for Developers被归为初级短期课程。相比之下,Deep Learning Specialization标注为中级,Fast and Efficient LLM Inference with vLLM同样是中级。仅从这些标注来看,不要求编程和数学的课程被归入初级,而需要亲手实现神经网络或对开源模型进行优化并部署的课程则被归入中级。值得注意的是,这份列表中没有任何一门课程被标注为高级。就连设计多个智能体的课程,或是从头训练语言模型的课程,也都没有正式的难度标注就被列了上去。公司内部围绕为何采用这种标注方式展开过怎样的讨论,现有资料无法确认。但仅从结果来看,这读起来像是一种结构:没有把难度一层层叠高来筛选学习者,而是只用初级和中级这两个阶段开门,再往上则留给有兴趣的人自行挑选进入。设计智能体或训练语言模型这些事所涉及

的技术本身固然复杂,但要上这些课程所必须通过的正式关卡,却没有另行设立。

这种难度标注实际遵循怎样的标准,在Deep Learning Specialization内部体现得更加具体。这门课程讲授的初始化方法都有各自的名字,Xavier初始化和He初始化。首次搭建神经网络时,每一层的权重从什么样的数字开始,会决定训练是顺利展开还是从一开始就卡住。如果权重填入的数值过小,信号每经过一层就会变淡,到后面几层几乎不剩什么信息;反过来如果数值过大,信号每经过一层就会膨胀,导致计算变得不稳定。Xavier初始化和He初始化正是在这两个极端之间,根据每层的规模计算出合适范围的方法。Dropout和批量归一化也处在类似的位置。Dropout在训练过程中随机关闭神经网络的部分神经元,使模型不至于只依赖特定路径;批量归一化则在训练过程中重新整理进入每一层的数值分布,以稳定学习速度。这些名字没有停留在只在论文中出现的术语层面,而是成为学习者在课程中亲自写代码时会碰到的对象,这一点让人能理解为何这门课程被标注为中级。

讲授vLLM的这门中级课程,性质也类似。这门课程教的是如何拿来已经做好的开源语言模型,对其进行优化使其响应更快,再把结果实际部署,并测量处理速度和资源使用量。制作新模型和让已有模型以服务能承受的速度运行,是两种不同的能力。后者很难成为不懂模型训练过程也能触及的领域,因为只有理解每次请求进来时模型会经过怎样的计算,才能判断在哪个环节能够节省时间。因此这门课程面向的与其说是第一次写Python代码的人,不如说是已经接触过模型的开发者。相比之下,在同一份面向从业者的最新课程列表中,用crewAI设计多个智能体的课程,或与Snowflake合作制作的多模态数据管道课程,都没有单独标注正式的难度。这些课程所涉及的技术虽然新颖而陌生,但由于是以用自然语言下达指令为核心,看起来并没有另设进入门槛。

这样看下来,DeepLearning.AI标注的初级和中级这两个阶段,遵循的是一种与技术新旧程度不同的标准。最近推出的课程未必就是难课程,反而是很早以前就存在的深度学习专项课程留在了更高的阶段。划分难度的标准不在于该课程所涉及的技术是何时问世的,而在于学习那项技术之前学习者需要预先具备什么样的知识。一门要求从头用代码搭建神经网络的课程,和一门只需用自然语言向已经做好的工具下达指令的课程,即便在同一时期推出,所要求的准备也不同。DeepLearning.AI的课程列表历经数年不断增加,但初级课程始终没有空缺过。AI for Everyone承担了这一角色,此后推出的生成式AI课程和Python入门课程也并排列在同一层次上。这说明即便技术在变化,给初学者留出的位置也始终有所准备。

这套两级标注体系,对想挑选一门课程的人来说起到了一枚小小的指南针的作用。它能减少从没编过程的人先打开标着中级的课程而不知所措的情况,也能减少已经训练过模型的人把时间浪费在初级课程上的情况。留言的学习者之所以说每门课程的路线图都很清晰,背后很可能也有这种标注在起作用,因为无论课程内容编排得多好,如果找不到适合自己的入口,这门课程就连开始都谈不上。DeepLearning.AI在不断制作新课程的同时,始终没有动摇这套简单的初级、中级两级标注,原因大概也在于此。

速度不同了,课程也不同了,学习者的处境也不同了。吴恩达没有把这些差异捆成一所长长的学校,而是把它分成了一扇扇短短的门。

# LandingAI与企业现场的人工智能

用教育的语言向世界授课的吴恩达再次创办了公司。这一次,他的学生不再是学生,而是企业。他要做的是把课堂上讲授的概念,转化为工厂、办公室、医院和律师事务所能够真正使用的形式。他创办的公司名为LandingAI,他以董事长兼首席执行官的身份领衔这家公司。

创办LandingAI之后,吴恩达先拿在手上的不是新技术,而是一份文件。他亲自撰写了名为《AI转型实战手册》的指导文件。这份文件开头有一句他早已琢磨已久的话。互联网进入世界的年代,无数公司只做了一个网站,就自称是互联网企业。吴恩达精准地指出了这种错觉。在购物网站上挂一个网页,并不能让它成为互联网企业。他把同样的逻辑原封不动地搬到了人工智能上。在一家普通的传统企业身上加装一项深度学习技术,并不能让这家公司变成人工智能企业。这两句话虽短,却清楚地表明了他在企业现场最警惕的问题。只改变表面的技术引入,和整个组织真正以不同方式运作的转型,是两回事。

这本手册涉及的范围并不止于某一项技术。用试点项目取得第一批成果、如何组建负责人工智能的团队、以何种战略收集和整理数据,以及在整个过程中如何在组织内外建立信任,这些内容都包含在内。他没有把AI转型的责任只留给技术部门一个人扛。他认为,从管理层的判断到一线人员的配合,再到处理数据的纪律,公司整体的体质都必须一同改变。购买新技术和用新技术改变工作方式,是两件规模不同的任务,他从不把这两件事混为一谈。

体现这一理念的正是如今的LandingAI。把公司的使命浓缩成一句话就是:让世界上的文档变得可计算。这个看似朴素的目标背后,是吴恩达长期抱持的一个问题。在尖端技术层出不穷的时代,真正困住大多数企业的,不是华丽的算法,而是枯燥且堆积如山的文档。合同和账单、保险单和病历、物流单据和法律文件,至今仍多是靠人一页页翻阅处理。吴恩达把这个由来已久的瓶颈,当作公司要正面攻克的课题。

如今的LandingAI并非由吴恩达一人执掌。他以董事会主席兼创始人的身份留守,实际经营则由首席执行官丹·马洛尼负责。英特尔资本和AI Fund旗下的MacroCap Capital等投资方为这家公司提供支持。他源自学术界的名字,如今已成为支撑投资人与经营团队共同运转的一个商业组织的核心支柱。

LandingAI面对的文档并不干净。实际企业处理的文档往往连成多页,格式因公司而异,不断出现偏离标准的例外情况。这与实验室里精心整理的样本数据截然不同,是杂乱无章的数据。这类文档由人工手动审核,长期以来在金融机构和保险公司被视为理所当然。LandingAI试图用能动式人工智能来取代这种人工审核。所谓「能动式」,指的是AI不是给出一个答案就停止,而是自主地持续完成阅读文档、做出判断、进入下一步这一连串工作。

实现这一目标的技术是能动式文档提取。这项技术并非单一的大整体,而是被拆分为多个细化功能。读取文档的解析、抽取所需信息的提取、搭建符合文档结构的框架的模式构建、判别文档种类的分类、按项目切分文档的分节,以及把长文档拆成便于处理的单位的分割,每个环节都以各自独立的API形式提供。企业可以只挑选适合自身业务的部分拼接使用。由此打造出的技术,延伸到了金融服务与保险、医疗健康、能源与公用事业、法律、物流等一切文档堆积的领域。

LandingAI将自己与其他人工智能创业公司区分开来的地方十分明确。那就是「原型之外,更要生产」这句话。只在演示视频里看起来运作得体的demo,和真正嵌入团队每日工作流、产出可衡量成果的系统,完全是两回事。吴恩达瞄准的是后者。他强调的第二条原则是「设计即信任」。自动化并不能仅凭准确

率一项就算完成。在风险高的行业——也就是一次错误判断就可能导致重大损失或法律责任的行业——AI给出的结果必须可验证、有据可依、可被审计。LandingAI没有止步于从文档中抽取数值,而是力图更进一步,把这些数值转化为能真正驱动下游系统和决策的可信输入。「从文档到系统」这句话正是这一方向的浓缩表达。

这一方向并未止于口号,成绩单便是证据。在衡量「阅读文档并回答问题」能力的DocVQA评测中,LandingAI的技术即便不依赖图像、仅凭文本也拿下了接近99.26分的成绩。Proxymo创始人P.K. Mishra在一项比较多种文档提取工具的基准测试中评价道,在全部四项评估指标上都取得高分的唯一工具,就是LandingAI的能动式文档提取。走出实验室之后,同样的评价再度出现。一家跻身财富100强的金融服务公司的数据与分析部门负责人表示,这套系统提供了完美的可追溯性。他补充说,正是这种可追溯性,才是让这项技术真正能够部署到监管严格的现实环境中的必要条件。

这一信任问题,在系统越复杂时表现得越鲜明。在打造需要穿梭于建筑法规知识图谱、跨越多张图纸和多个专业领域进行推理的「图纸审查智能体」时,开发者们碰到的墙壁最终落在同一个地方。如果无法信任从页面中提取出的信息,那么在其之上堆叠再精巧的推理也毫无意义。无论算法多么华丽,只要输入值出现动摇,整个系统就会随之动摇。这正是LandingAI在文档提取这一件事上倾注如此心力的原因。吴恩达选择的,不是引人注目的算法,而是不太显眼的基础——也就是数据的可信度。

LandingAI把这项技术开放出来,让人无需注册信用卡,几分钟内即可试用。只要上传一份复杂的文档,任何人都能亲眼看到它被转化为结构化数据的过程。曾在课堂上教授概念的他,这一次让企业的负责人亲手触摸产出的结果。教学的方式变了,但降低陌生技术门槛的态度一以贯之。

曾在大型科技公司里做扩大神经网络规模的实验、又跨出校园围墙把课程传播到全世界的吴恩达,如今正面对企业抽屉里沉睡多时的成堆纸张。比起耀眼的研究成果,他选择了枯燥的实务问题,这一点体现出LandingAI折射出的一条轨迹——他从学界转向产业,又转到产业中最不起眼的角落。在他随后创立的组织中,不论是扶持创业的AI Fund,还是他跻身董事会的亚马逊等场合,这种态度依然延续。吴恩达优先要立起来的,不是宏大而华丽的承诺,而是一个真正能运转的系统。

再仔细看看LandingAI处理的自动化范围,就能看出这家公司从一开始就没有只瞄准某一个行业。吴恩达选定的是「文档」这个对象本身。无论哪个行业,几乎都会积累合同、账单、申请表和凭证。他没有选择成为某一特定行业的专家,而是选择深挖横跨各行业的一个共同瓶颈。这一选择,与他在Google Brain埋头于扩大神经网络规模实验的那段岁月,以及在Coursera向世界多国学生同时开放同一门课程的那段岁月,都是一脉相承的。无论哪个阶段,他做事的方式都不是深挖某个狭窄的专业领域,而是找出多处同时受阻的共同墙壁。

能动式文档提取被拆分为多个小功能来提供,这一点也与这种思维方式相通。解析与提取、模式构建与分类、分节与分割这六个组成部分各自以独立的API开放,企业可以从中只挑选自己需要的部分接入系统。保险公司可以只取用文档分类功能,律师事务所可以只挑选把长合同按条款拆分的功能。这种组合方式,与销售技术的公司常常陷入的陷阱——强迫客户整体购买一个庞大的整体方案——性质不同。吴恩达以每家企业所处情况各不相同为前提,把工具设计得可以细分拆解,以吸收这种差异。

这种做法之所以能够实现,背后是他长期强调的一条原则。那就是:决定系统实际性能的,与其说是好的算法,不如说是好的数据。学术界的基准测试大多是在整理干净的样本上一较高下。然而,现实企业抽屉

里的文档扫描质量参差不齐,夹杂手写内容,同一项目在不同公司也叫法各异。LandingAI没有只强调实验室基准分数,而是反复强调处理实际运营环境中脏乱数据的能力,原因就在于此。仅凭准确率这一个数字无法说明这种差异。系统在遇到陌生格式和例外情况时能否不崩溃、依然给出结果,在实际现场是重要得多的标尺。

「设计即信任」这条原则在这里再次显出意义。吴恩达没有把自动化只当作准确率的问题。他认为,结果产生的路径必须能让人重新核实。可验证意味着有办法重新核查结果,有据可依意味着结果来自文档的实际内容,而不是系统凭空捏造的推测。可审计意味着即便日后有人也能追溯这一判断为何得出。在金融和保险这类监管细密的行业,首先要考量的不是某个结果是否准确,而是能否说明为什么可以相信这个结果。LandingAI提出的三项条件,正是对这些行业长期以来的要求给出的正面回应。

这条原则并未停留在抽象的口号上,这一点在一家跻身财富100强的金融服务公司的证词中再次得到确认。该公司数据与分析部门负责人提到的「完美的可追溯性」,意味着不仅结果的准确率,能够回溯结果产生过程这一点,才是决定实际部署与否的标准。监管越严格的环境,越是由系统给出的答案所经路径是否透明来决定是否采用,而非答案本身。LandingAI把公司的资源集中投入到满足这一条件上。

围绕图纸审查智能体的案例,从另一个角度展示了这条原则为何必要。建筑法规是一个多张图纸与多个专业领域规定相互交织的领域。管道图纸、电气图纸、结构图纸各自遵循不同规定,却要在同一栋建筑中彼此契合。开发者们在打造这类系统时碰到的墙壁,不是缺少华丽的推理算法。而是一个简单的事实:从页面中提取的某条信息一旦出错,不管在其之上堆砌多么精巧的推理,整体结果都会随之动摇。这是任何设计系统的人迟早都会遇到的教训。比起引人注目的上层算法,不那么显眼的底层数据质量,决定了整个系统的上限。吴恩达之所以把公司的力量投入到文档提取这一表面看起来并不华丽的课题上,看来正是因为他在实务中反复确认了这一教训。

LandingAI瞄准的行业清单——金融服务与保险、医疗健康、能源、法律与物流——乍看彼此相距甚远。然而这些行业有一个共同点。那就是纸质文档和扫描件长年累积,而处理这些文档的工作至今仍高度依赖人的手和眼。律师事务所审阅合同、保险公司审核理赔、物流公司处理报关单据,都是需要熟练人员逐页阅读文档的工作。这类工作既不华丽,也很少登上媒体版面,却实实在在地左右着企业的人力成本和处理速度。吴恩达精准地找到了这个不显眼的位置。

这一选择也与他在Coursera和DeepLearning.AI展现出的态度一脉相承。他总是找到最多人受阻的地方,并在那里开出一扇门。在线教育降低了大学课堂的物理门槛,DeepLearning.AI把只有专业研究者才懂的概念转译成从业者的语言。而在LandingAI,那道门槛不是技术本身,而是数据的形态。把被困在「文档」这一形态里的信息,转化为系统能够处理的结构——这正是LandingAI反复提及的「让其变得可计算」这句话的实际内容。

吴恩达通过这家公司展现出的态度,最终归结为一点。那就是:添加一项新技术,和让组织真正以不同方式运作,是两回事。LandingAI之所以开放了无需注册信用卡即可试用一份文档的工具,也与这条原则不无关系。企业负责人亲自把自己的文档放进去查看结果的体验,比口头说明的信任更快建立起确信。吴恩达早在Coursera时代就知道,建立这种确信最短的路径就是让人亲自动手一试,在LandingAI他把同样的方式原封不动地搬了过来。

LandingAI所涉及的各行业监管强度各不相同,而这种监管强度甚至决定了对待文档的态度。金融监管机构要求,即便几年之后也要能够回溯审查依据;保险监管机构要求说明理赔核准与拒赔的理由出自文档的哪个部分。律师事务所处理的合同,一处条款解释的分歧就可能演变成诉讼;医疗机构的记录则直接关系到患者安全。在这些行业里,系统能否快速给出一个答案是次要项目。能否说明这个答案从何而来才是首要的。LandingAI之所以并列提出可验证性与有据可依的可审计性这三项条件,背后正是这些不同行业要求的叠加。

「能动式」这个名字的由来,在这里也值得再次审视。过去的文档处理自动化,常见的做法是只识别符合固定模板的文档,超出模板的文档就转交给人处理。LandingAI强调的能动式方式,把阅读文档这一过程本身拆分成多个阶段,每个阶段都做出判断再进入下一阶段。在解析阶段读取文档的物理结构,在分类阶段判别这份文档属于哪一种类,在模式构建阶段搭建适合该文档种类的框架,在提取阶段抽取符合该框架的数值。这一流程与人审阅文档时所走的步骤相似。先浏览文档种类,确定要看哪些项目,再找出所需数值并誊录下来。LandingAI没有把这整套流程装进一个整体箱子里,而是按阶段逐一开放,让企业可以在与自身审阅流程契合的节点插入这套系统。

之所以需要这种方式,是因为文档处理这项工作原本就要求人的判断。一张账单是否核准,并不是简单的文字识别就能了结的问题。需要同时核查账单金额是否与合同条款相符、随附的凭证是否合乎规定、与过往记录相比是否存在异常。LandingAI之所以在提取之后另外配置分类、分节、分割功能,正是为了把这类附加判断纳入系统之中。这不只是从一份文档中抽取数值,而是连同数值所处的语境一并整理,让下一个系统能够直接拿来使用。

DocVQA中拿下的99分左右的成绩,说明这项技术能够准确读取文本,但企业并不会仅凭这一个数字就决定采用与否。财富100强企业负责人提到的「完美的可追溯性」这一表述,点明了准确率这把尺子和信任这把尺子处于不同的层面。准确率说的是系统给出正确答案的比例,可追溯性说的是能否回溯这个答案的出处。在受监管行业里,如果没有后者,前者再高也难以把系统真正投入实际业务。因为当审计人员或监管机构日后追问这一判断是如何得出时,必须能够展示这个数值来自文档的哪一句话。LandingAI提出的「从文档到系统」这句话,可以理解为在每一个提取出的数值上都附上其出处再传递下去的意思。

打造图纸审查智能体的开发者们碰到的墙壁,也是同一个故事的另一面。建筑图纸拆分成多张,管道、电气、结构各自遵循不同规定,却要在同一栋建筑中相互契合。在这样的系统里,无论推理阶段多么精巧,只要从图纸中一开始就读错了某个数字,建立在其上的判断就会全部导向错误的结论。开发者们确认的事实终究是:决定系统上限的不是推理的华丽程度,而是提取的准确程度。这正是LandingAI把公司资源集中投向文档提取这一表面并不华丽的课题的原因所在。

再次审视LandingAI瞄准的行业清单,金融与保险、医疗健康、能源、法律与物流这六个项目乍看彼此相距甚远。然而它们都是纸质文档和扫描件长年累积的行业,处理这些文档的工作至今仍高度依赖人的手和眼,这是它们的共同点。律师事务所审阅合同、保险公司审核理赔、物流公司处理报关单据,都是需要熟练人员逐页阅读文档的工作。这类工作虽不显眼,却实实在在地左右着企业的人力成本和处理速度。吴恩达精准地找到了这个不显眼的位置。

这一选择也与他在Coursera和DeepLearning.AI展现出的态度一脉相承。他总是找到最多人受阻的地方,并在那里开出一扇门。在线教育降低了大学课堂的物理门槛,DeepLearning.AI把只有专业研究者才懂

的概念转译成从业者的语言。而在LandingAI,那道门槛不是技术本身,而是数据的形态。把被困在「文档」这一形态里的信息,转化为系统能够处理的结构——这正是LandingAI反复提及的「让其变得可计算」这句话的实际内容。

吴恩达通过这家公司展现出的态度,最终归结为一点。那就是:添加一项新技术,和让组织真正以不同方式运作,是两回事。LandingAI之所以开放了无需注册信用卡即可试用一份文档的工具,也与这条原则不无关系。企业负责人亲自把自己的文档放进去查看结果的体验,比口头说明的信任更快建立起确信。吴恩达早在Coursera时代就知道,建立这种确信最短的路径就是让人亲自动手一试,在LandingAI他把同样的方式原封不动地搬了过来。

LandingAI围绕文档这件事持续深耕的同时,吴恩达赋予这家公司的一条准则也格外醒目。那就是「原型之外,更要生产」这句话。他之所以特意把这句话放在前面,与AI创业公司行业常见的现象不无关系。不少系统在演示环节运转顺畅,一旦投入公司真实的日常业务就停摆。一名负责人在屏幕前演示时能够顺利给出答案,可当数百人每天涌入不同格式的文档时,系统却可能崩溃。吴恩达把这道鸿沟当作公司的第一判断标准。也就是说,只有在整个团队和工作流程中都能产出可衡量成果的系统,才被认定为产品。

这条标准也与他在学界工作时期的习惯一脉相承。研究论文只需证明一种新方法比现有方法取得更好的分数即可。但进入企业的系统不会止于一个分数。它必须经受住今天早晨新流入的陌生格式文档、扫描质量差的文档,甚至负责人手写批注的文档的考验。吴恩达在LandingAI特意把基准测试分数和实际企业的证词并列摆出,原因在于这两种验证说明的是不同的事情。基准测试展示的是系统有多准确,实际部署案例展示的则是这种准确能否在陌生环境中依然站得住。

从LandingAI所涉行业的广度来看,也能看出这家公司并非把单一技术换个名字到处贩卖。金融与保险处理理赔与核准,医疗健康处理诊疗记录与处方,能源与公用事业处理设备巡检报告。法律处理合同,物流处理报关单据。这六个领域的文档形态各不相同,监管强度也各不相同。尽管如此,LandingAI之所以能用一项技术涵盖所有这些领域,是因为吴恩达从一开始瞄准的不是某个行业的专属语言,而是「文档」这一共同形态。这一决定与其说是他事先调研多个行业后定下的战略,不如说更像是源自一个简单的观察——文档这个对象本身不分行业地在不断累积。

能动式文档提取被拆分为六个API提供,这一事实也与这种观察相呼应。解析与提取、模式构建与分类、分节与分割这些组成部分,各自都可以独立调用。如果一家保险公司只需要对账单进行分类的功能,拿来这一项功能即可。如果一家律师事务所只需要把长合同按条款拆分的功能,接入那一块就够了。这种组合方式,前提是承认每家企业已经具备的系统各不相同。它不是要求企业把整套系统按LandingAI的方式推倒重来,而是LandingAI这一方去契合企业既有的工作方式。

这种灵活的组合背后,是吴恩达长期以来坚持的一个想法。那就是:决定系统实际性能的,与其说是一个好的算法,不如说是好的数据。学术界的基准测试大多是在整理干净的样本上一较高下,而企业抽屉里的文档扫描质量参差不齐,夹杂手写内容,同一项目在不同公司也叫法各异。LandingAI没有只强调实验室分数,而是反复强调处理杂乱实际数据的能力,原因就在于此。仅凭准确率这一个数字无法说明这种差异。系统在遇到陌生格式和例外情况时能否不崩溃、依然给出答案,在实际现场是分量重得多的标尺。

「设计即信任」这条原则在此再次显出意义。吴恩达没有把自动化只当作准确率的问题。他认为,结果产生的路径必须能让人重新核实。可验证意味着有办法重新核查结果,有据可依意味着结果来自文档的实

际内容,而不是系统凭空捏造的推测。可审计意味着即便日后有人也能追溯这一判断为何得出。在金融和保险这类监管细密的行业,首先要考量的不是某个结果是否准确,而是能否说明为什么可以相信这个结果。

一家跻身财富100强的金融服务公司的证词表明,这条原则并未停留在口号上。该公司数据与分析部门负责人提到的「完美的可追溯性」,意味着不仅结果的准确率,能够回溯结果产生过程这一点,才是决定实际部署与否的标准。监管越严格的环境,越是由系统给出的答案所经路径是否透明来决定是否采用,而非答案本身。

围绕图纸审查智能体的案例,从另一个角度展示了这条原则为何必要。建筑法规是一个多张图纸与多个专业领域规定相互交织的领域。管道图纸、电气图纸、结构图纸各自遵循不同规定,却要在同一栋建筑中彼此契合。开发者们在打造这类系统时碰到的墙壁,不是缺少精巧的推理算法。而是一个简单的事实:从页面中提取的某条信息一旦出错,不管在其之上堆砌多么精巧的推理,整体结果都会随之动摇。比起引人注目的上层算法,不那么显眼的底层数据质量,决定了整个系统的上限。吴恩达之所以把公司的力量投入到文档提取这一表面看起来并不华丽的课题上,看来正是因为他在实务中反复确认了这一教训。

LandingAI开放无需注册信用卡即可试用一份文档的工具,这一决定同样与这条原则不无关系。企业负责人亲自把自己的文档放进去查看结果的体验,比口头说明的信任更快建立起确信。曾在Coursera时代把在线课程向世界各地学生开放的吴恩达,当年就是先展示课程资料,再让学生自己判断是否报名。在LandingAI,同样的顺序再次重演。先展示结果,信任交由看到结果的人自行建立。虽然从课堂转到了公司,但在陌生对象面前先伸出手的方式,并没有太大改变。

# AI Fund、亚马逊与创业生态

在Landing AI,吴恩达确认的是改变一家企业这件事的分量。即便是把一家制造工厂里的一个文件处理部门转变为AI驱动,也需要漫长的时间和精细的工作。这段经历自然而然地引出了下一个问题:未来会有更多想用AI改变世界的创业者出现,他们缺少的是什么?吴恩达给出的答案不仅仅是资金,还包括验证创意的眼光、组建早期团队的判断力,以及把技术转化为实际业务的实务知识。把这个答案落实为组织形式的结果,就是AI Fund。

AI Fund选择了一条不同于通常那种出资换取股权的风险投资公司的道路。他们自称为风投工作室。风险投资公司挑选已经成立的公司进行投资,而风投工作室则从创意阶段就与创业者一起创建公司。吴恩达以执行合伙人的身份领导这个组织,坚持与创业者并肩走过发掘创意、验证市场、建立可持续业务的全过程。AI Fund关注的焦点并非公司已经站稳脚跟之后,而是连商业计划书都还没有出现的极早期阶段。他们判断,在这个阶段提供帮助,才能创造出最大的差异。

为这一判断出资的机构,在硅谷风投圈里并不陌生。New Enterprise Associates、红杉资本、软银集团都作为有限合伙人参与其中。此后AI Fund又组建了用于联合创办新AI初创企业的第二期基金,即所谓的Fund II,规模达1.9亿美元,并以超额认购的方式收官,超过了目标金额。这意味着大型机构投资者争相注入资金。这个组织的底层逻辑,建立在吴恩达自谷歌大脑时代起反复强调的一句话之上:AI是新的电力。正如电力普及到各个产业、催生了无数新产品和新服务一样,人们相信AI也将为数万个新应用提供动力。AI Fund正是把这种信念落实到每一家初创企业层面的尝试。

为什么选择风投工作室而不是风险投资?这一选择背后可能存在的个人思考,或与其他投资模式的比较过程,在公开资料中未见记载。不过,由此产生的组织结构和运作方式却是清晰的。

翻看AI Fund迄今培育的公司名单,可以看出他们瞄准的领域并不狭窄。其中有打造实验管理工具的RapidFire AI,有帮助企业安全管理AI工具使用的Olakai,还有以让每个孩子都拥有基本识字能力为目标的Esteem。此外还有帮助投资者发现隐藏机遇与风险的Octagon,以及帮助容易在组织中失去存在感的中层管理者树立信心、带领团队的Ripple,也在这份名单之列。在飓风等灾难现场执行人命救援任务的自主无人机平台Skyfire AI,证明了AI Fund并未止步于消费者服务。此外还有通过语音直译语音降低语言障碍的SpeechLab,为销售与市场团队打造客户情报的Sunrise AI,借助AI职业指导把职务转化为职业热情的Worka,以及为大型企业规划AI生产力路径的WorkHelix。这份名单横跨教育、灾难应对与企业软件等多个领域。

AI Fund的活动范围也扩展到了美国之外。他们通过亚洲枢纽宣布进军台湾,并与三菱商事携手开展AI创新工作。与能源企业AES缔结了在六个月内共同创办AI能源初创企业的合作关系。这种区域扩张与分行业合作是否源于吴恩达个人的某种契机,资料中并未显示。可以确认的只是,AI Fund已经把触角伸出美国消费市场,延伸到了亚洲和能源产业。

在培育初创企业的同时,吴恩达也着手改造那些已经站稳脚跟的大型企业。如果说Landing AI帮助的是制造业现场引入AI,那么AI Aspire则是面向更广泛大企业的咨询组织。吴恩达与同事Kirsty Tan共同创立了这家公司,并以执行合伙人身份参与其中。他在一次采访中表达的问题意识很明确:要真正改

变AI应用的格局,仅仅吸引开发者或个人消费者是不够的,必须撬动大型企业。这句话里包含着这样的判断——一款个人聊天机器人无论多么普及,也无法改变整个产业的生产率曲线。

为实现这一目标,AI Aspire与全球知名管理咨询公司贝恩公司的团队缔结了合作关系。领导这次合作的是贝恩方面的克里斯托夫,两家机构直接面向大企业高管提供咨询。他们瞄准的,是超越企业围绕AI展开的小规模试验。让一个试点项目取得成功,和把这种成功推广到整个组织、真正改变结果,是两回事。AI Aspire制定的战略目标,正是要创造这种可规模化、具有变革性的结果。吴恩达在做出这一决定之前经历过怎样的思考,以及与大型咨询公司合作在内部是否存在异议,公开资料中都无法确认。

2024年4月9日,吴恩达的履历上又添了一个新头衔。他加入了全球最大电子商务企业兼云计算企业亚马逊的董事会,成为董事。这一职位是因为在董事会任职十年的朱迪·麦格拉思当年在年度股东大会上决定不再谋求连任而空出的。亚马逊向吴恩达提出邀请、吴恩达接受邀请之间经过了怎样的沟通或个人交情,外界并不知晓。可以确认的只是结果——他成为了董事会的一员。

亚马逊发布的官方声明说明了这次任命的理由。公司将人工智能、尤其是生成式人工智能定位为这个时代最具变革性的技术之一,并说明需要吴恩达的经验来帮助董事会层面理解这项技术带来的机遇与挑战。声明具体列举了吴恩达的履历:他撰写过200多篇机器学习和机器人学论文,是谷歌大脑的创始负责人,曾在百度担任首席科学家,也是Coursera和Landing AI的创始人。学术界积累的研究经历与在多家企业积累的经营经验被一并提及,说明亚马逊对他的期待并不止于单纯的技术顾问角色。董事会表示,他的洞见将有助于董事会整体形成关于AI将给社会和商业带来何种变化的看法。

如果要用一个词概括吴恩达在多个组织中所扮演的角色,那就是导师。他通过AI Fund尝试提供超越投资者与创业者关系的支持。他所表达的态度是:自己敬重那些坚信自己所相信的事、并为把它变为现实而努力的企业家,而帮助这些人在自己的领导下组建合适的团队,是自己看重的核心价值。这番话可以理解为,吴恩达把自己定位为不是单纯出资的一方,而是与创业者共同分享技术、组织运营和商业判断力的一方。在这一辅导过程中他与创业者之间可能经历过的分歧或情感交流,公开资料中并未涉及。

这种导师身份具体体现在AI Fund运营的洞见频道和博客上。这里积累着早期创业者可以立即拿来使用的实务指南。频道上刊登了面向早期AI初创企业的增长营销方法、打造能从投资者手中获得资金的融资演示文稿的方法、招聘公司第一位销售人员的方法,以及关于初创企业讲述自身故事的三大支柱的文章。也涉及组建技术团队的问题:创业者在招聘技术团队之前应该了解的事项、当团队所有成员都能写代码时组织会呈现出怎样的面貌的展望,以及构建大语言模型技术栈时面临的各种取舍分析,都是其中的例子。还有关于行业现场的文章:通过风投工作室与企业合作从AI转型中获得的洞见,以及传统向量搜索在企业环境中失效的原因和知识图谱这一替代方案的文章。这个频道上的文章一律免费公开,即便不是吴恩达所在组织直接投资的创业者也能够阅读。

吴恩达并不只通过文字分享这些知识。他还曾邀请Pandora创始人蒂姆·韦斯特格伦,主持了一场以坚持和领导力为主题的大师课对话。这是一次尝试,把长期经营公司的创业者的经验直接传递给下一代创业者。AI Fund通过名为驻场工程师的项目招募申请者,并在山景城举办建造者活动,与下一代AI创业者面对面交流。在这样的场合,吴恩达提出的目标可以概括为一句话:培育推动人类进步的伟大企业。在这类面对面活动中,吴恩达向参与者给出了怎样的个人建议,资料中并未留存。

作为导师,吴恩达的活动并未止步于企业现场和创业生态,而是延伸到了政策领域。他向美国参议院设立的AI Insight Forum提交了书面声明,就国家层面的AI政策和监管方向发表了意见。他作为研究者、创业者和企业董事所积累的经验,成为政策讨论的素材。这种广泛的活动也反映在外界的评价上——2023年《时代》杂志评选的AI领域最具影响力100人榜单上,出现了他的名字。

本章所呈现的吴恩达的轨迹,与此前的行迹有着不同的色调。如果说在谷歌大脑他从事的是大规模研究,在Coursera和DeepLearning.AI从事的是教育,在Landing AI处理的是单个企业的转型,那么在AI Fund、AI Aspire和亚马逊董事会,他动用的则是资本、董事会和政策这些技术之外的工具。此前那种把实验室验证过的想法搬进课堂、再把课堂上遇到的人的需求转化为企业现场问题的模式,这一次获得了资本和组织这一新的通道。不过这条通道究竟带来了怎样的实际成果——投资组合初创企业的存活率,或AI Aspire提供咨询的企业取得的具体业绩——公开资料并未提供确认依据。可以确认的只有组织的结构、参与者名单,以及他们所提出的目标。明确这一局限,有助于不夸大地理解吴恩达这一时期的活动。

出资时点的差异,才是区分风险投资与风投工作室的真正界线。传统风险投资出现在创业者已经创办公司、甚至做出了早期产品之后。投资者审阅商业计划书和早期指标,如果满意就以股权换取资金。确定公司名称、寻找联合创始人、见第一位客户,这些事往往已经由创业者独自完成。AI Fund选择的方式把这个顺序提前了。从想法连一页文档都还没成形、创业者只在笔记本上写下假设级构思的阶段起,这个组织就介入进来,与创业者一起搭建公司的框架。在这种结构下,投资者和创业者的界限变得模糊。AI Fund的人员会一同做早期市场调研,一同起草第一份招聘启事,与创业者并肩梳理产品第一版需要解决的问题。吴恩达以执行合伙人这一头衔领导这个组织,也可以从这个脉络来理解。这个位置与其说是投资评审委员的位置,不如说更接近于同时参与多家公司创办过程的联合创始人的位置。

仅凭Fund II超额认购、超过目标金额收官这一件事,就有值得推敲之处。在风投行业,超额认购是指投资者愿意投入的资金超过基金计划募集的金额,以至于运营方不得不退回部分资金的情形。这意味着,资金短缺的不是初创企业,反倒是基金处在挑选资金的一方。这种结果并不常见,因为风投工作室这一模式本身相对于传统基金而言还比较新,可供验证业绩的数据也相对较少。New Enterprise Associates、红杉资本、软银集团这类有限合伙人——也就是向基金出资、把运营交给他人的机构投资者——向这一模式注入资金,这一事实在业内被视为一种信号。这三家机构各自以不同的方式在硅谷长期占有一席之地。这些机构参与验证,可以理解为不仅是对吴恩达个人声誉的认可,更是对他所创立组织的运作方式本身被认定为一种可重复的结构。

再次审视投资组合公司的名单,会发现一个规律。识字教育、灾难救援无人机、语音翻译、销售支持软件、面向中层管理者的辅导工具、投资风险分析——这些公司行业不同、客户不同、技术用途也不同。这种多样性并非偶然,而是与AI Fund所设定的前提相呼应。如果一项技术是像电力那样在全社会普及的通用技术,那么它就应当在看似互不相关的多个产业中同时催生出新的用途,而不只是在一个产业里。电力最初普及时,工厂的生产线、家庭的照明、医院的医疗设备,各自以不同的方式接纳了电力。只改变了一个产业的技术,规模就不算大。AI Fund之所以跨多个行业创办公司,某种程度上也是在检验这个组织自身设立的假设。不过,这种投资组合布局究竟是从一开始就设计好的战略,还是随着不断涌入的创业者想法而自然扩展的结果,公开资料无法判定。

AI Aspire瞄准的问题,与业内通常所说的试点炼狱现象相关。许多大企业尝试引入AI时会经历一条常见的路径。某个部门用少量预算启动一个试点项目,这个项目在会议室里看起来很成功,负责人也做好了推

进下一阶段的准备。但这个试点项目很少能真正推广到整个公司。原因往往不在于技术本身的不足,而在于组织结构。一个部门行得通的方式,可能与另一个部门的数据体系或业务流程不兼容,试点阶段被默许放过的例外情况,到了全公司推广阶段就变成了绊脚石。如果说Landing AI专注于改变单个工厂、单个部门,那么AI Aspire的定位则是超越试点阶段、推广到整个组织的阶段。它与贝恩公司这样的大型管理咨询公司携手,也可以从这一点来理解。懂AI技术,和同时改变数千名员工组织的决策结构、预算分配方式与人事考核标准,是两种不同性质的工作。后者更接近长期处理经营组织问题的咨询公司的领域,前者则是吴恩达长期积累的领域。两个组织的结合,可以视为把这两者纳入同一套咨询之中的尝试。

加入亚马逊董事会,与前面提到的两个组织性质不同。在AI Fund和AI Aspire,吴恩达处在直接参与组织日常运营的位置。董事会则不同。上市公司的董事会并非直接处理公司日常业务的岗位,而是监督管理层做出的重大决策、代表股东批准或叫停公司大方向的岗位。董事更接近于对股东负有诚信义务的监督者,而不是为公司工作的员工。进入2020年代后,随着生成式人工智能成为多个行业的热门议题,不止亚马逊一家,多家上市公司都出现了在董事会成员中安排深谙AI的人物的趋势。董事会做出的决策——比如向哪项业务投入资本、可以承担哪些风险、管理层提出的AI相关业务计划是否现实——都需要董事们自己对这项技术的局限和可能性有一定程度的把握才能判断。亚马逊在引进吴恩达时发表的声明中提到,需要他的经验来帮助董事会层面理解AI带来的机遇与挑战,这句话可以放在这个背景下理解。声明中一并提到他作为研究者撰写论文、在谷歌大脑和百度运营组织、创办Coursera和Landing AI的履历,也是出于同样的脉络。亚马逊需要的不是只懂技术本身的专家,而是一个亲身经历过技术转化为实际业务时会遇到的多层面问题的人物。

董事会的工作,表面上呈现出来的活动并不多。董事会会议不公开,会议上讨论的内容、吴恩达提出的问题、他反对或赞成的议案,都无法通过公开资料确认。他作为董事实际产生了怎样的影响,目前没有可供判断的依据。不过,他担任这一职位这件事本身,说明他的活动范围已经超越初创企业或中型企业,扩展到了全球屈指可数的大型科技企业治理结构之中。

把吴恩达作为导师所做的事与前面几章并列来看,能看到一条贯穿始终的线索。在Coursera和DeepLearning.AI,他让想学习人工智能的人不必跨过课堂的门槛。在AI Fund的洞见频道,他让想创业的人不必支付昂贵的咨询费用。撰写融资演示文稿的方法、招聘第一位销售人员的方法、处理大语言模型时遇到的技术取舍等文章免费公开,不仅AI Fund直接投资的创业者可以阅读,任何在这之外的人也都能够阅读。用讲课分享知识的方式这次变成了实务指南这种形式,但降低门槛的方向大体没有改变。邀请Pandora创始人蒂姆・韦斯特格伦举行对话的做法,也属于这一脉络。这是一个把长期经营公司的创业者所经历的摸索,以对话而非课堂的形式传递给下一代的场合。

名为驻场工程师的项目,也是这种导师身份的延伸。风投工作室常运营的这类项目,做法是把还未加入特定公司的工程师留在组织内部,让他们往返于多个早期项目之间,一起解决技术问题。这是一种让工程师在公司成立之前就与组织建立关系的机制。在山景城举办的建造者活动,与其说是招揽投资的场合,不如说更接近于让尚未创办公司的人们相互结识、试验想法的场合。在这样的场合,吴恩达提出的目标——培育推动人类进步的伟大企业——与他长期反复使用的语言相呼应。不过,这句话具体是在什么情况下、对哪位创业者说出的,公开资料中并未确认。

向美国参议院AI Insight Forum提交书面声明一事,也值得关注。这个论坛是2023年多次召开的非公开简报性质的会议,召集科技企业高管、研究者和民间团体人士,共同讨论人工智能监管的方向。研究者出身的人士在这类场合提交书面意见并不罕见,但吴恩达被邀请的背后,很可能与他横跨学术界、产业界和教育现场的履历有关。他被选为政策制定者可以参考的声音这一事实,意味着他的发言已经超越单个企业或单门课程,触及到了国家层面的讨论场合。2023年《时代》杂志将他列入AI领域有影响力人物100人榜单,也是这种广泛活动在外界如何被解读的一个指标。

本章所涉及的三个组织——AI Fund、AI Aspire和亚马逊董事会——各自在不同层面运作。AI Fund是创办尚未诞生的公司的场所,AI Aspire是改变已经站稳脚跟的大企业内部结构的场所,亚马逊董事会则是监督一家全球规模企业重大决策的场所。在这三个岗位上,吴恩达都不是亲手打造技术的人,而是帮助技术通过组织、资本和政策这些形式扩散到世界的人。这些活动究竟带来了多大的实际变化,投资组合公司有多少存活下来,他在董事会上就哪些议案发过声,以目前公开的资料都无法得知。可以确认的,只有他同时担任这三个职位这一事实,以及这些职位所瞄准的目标。

要完整理解吴恩达在AI Fund选择的方式,需要把它与创业生态中已经存在的另外两种方式加以比较。一种是加速器。在短时间内把多个团队聚集在一起,按照既定日程给予指导,同时投入少量资金、只换取很小比例的股权。另一种是传统风险投资。它在创业者已经创办公司、做出一定形态的产品之后才出现,出资换取股权,但不深入参与公司运营。AI Fund选择的风投工作室模式与这两者都不同。它从公司诞生之前就参与进来,一起打磨想法、组建团队、确定产品方向。这种差异也直接体现在股权结构上。加速器获得的股权比例不大,传统风险投资获得的股权也是以创业者已经积累的价值为基准计算的。风投工作室因为从公司还不存在的时候就开始共同工作,往往采取创业者与工作室更加对等分配股权的结构。在这种结构下,创业者手中掌握的股权比例可能会减少,但创办公司所需的时间和风险也相应减少。

值得留意的是,吴恩达在AI Fund和在AI Aspire使用的头衔并不相同。在AI Fund,他使用的是执行合伙人这一头衔。这个头衔指的是负责运营投资基金的人,通常与基金所投资公司的股权价值上涨后分得部分收益的结构相配合。在AI Aspire,他使用的是管理合伙人这一头衔。这是收取咨询费、提供服务的专业服务型组织常用的头衔。两个头衔的差异,说明这两个组织获取收益的方式不同。AI Fund通过创办公司、让其价值上涨来获利,AI Aspire则通过向大企业提供咨询获得报酬。同一个人,正在以两种不同的方式参与人工智能的推广。

为AI Fund出资的有限合伙人的构成,也有必要再次审视。New Enterprise Associates和红杉资本本身就是直接投资初创企业的风险投资机构。这类机构在自己运营的基金之外,以有限合伙人身份向他人的基金出资的情况并不罕见。规模较大的风险投资机构有时会选择把资金托付给擅长在极早期与创业者共同创办公司的组织,再分享其成果。AI Fund能够吸引到资金,背后很可能存在这样的考量。不过,各机构经过怎样的内部审查才做出投资决定,公开资料并未提供确认依据。

名为驻场工程师的项目,也属于风投工作室行业中广泛使用的机制。把尚未隶属于特定公司的工程师留在工作室内部,让他们往返于多个早期项目之间解决技术问题,这一做法是为了解决工作室创办公司时经常遇到的一个难题。即便有好的想法,如果没有能把想法转化为代码的工程师,公司也无法启动。工作室提前储备工程师,就能加快想法与人才配对的速度。AI Fund运营这一项目的事实,说明这个组织不仅仅止步于出资,也把招揽人才视为自己的职责之一。

在董事会中安排懂人工智能的人物这一趋势,并不止于亚马逊一家。进入2020年代中期后,多家上市公司为了辅助管理层的判断,纷纷在董事会成员中安排深谙人工智能技术的人。在这一趋势之下,吴恩达加入亚马逊董事会一事,不仅是他个人的事件,也可以视为科技企业治理结构整体正在经历的变化的一个例证。由于董事会是批准或叫停公司业务计划的场所,董事们是否具备判断某项特定技术可行性与风险的能力,对股东而言也是重要的问题。亚马逊在声明中逐一列举他的履历,背后也正是出于这种需要。

AI Fund在洞见频道上发布的文章免费公开这一点,与吴恩达在Coursera和DeepLearning.AI一贯坚持的方式相呼应。不过受众范围有所不同。Coursera的课程面向所有想学习人工智能的人,而AI Fund的实务指南只面向已经下定决心创业或正在筹备创业的人。随着受众收窄,内容也变得更加具体。讲解如何撰写融资演示文稿或如何招聘第一位销售人员的文章,处理的是与那种要从人工智能是什么讲起的入门课程不同层次的知识。这种差异说明,吴恩达一直在根据受众调整降低教育门槛的方式。

把本章所涉及的三种角色并列来看,可以看出吴恩达所做工作的性质发生了变化。在谷歌大脑,他亲自设计神经网络;在Coursera,他亲自录制课程。而在AI Fund、AI Aspire和亚马逊董事会,他不再亲手写代码或设计产品。他积累的知识和经验,只是被用作帮助他人做出判断的素材。这种变化与其说是影响力减小,不如说是影响力扩散的渠道增多了。一个人能够亲手创造的东西是有限的,但把判断所需的素材交给多位创业者、多家企业、多位政策负责人这件事,受到的限制相对更少。AI Fund培育的多家公司、AI Aspire提供咨询的多家大企业、亚马逊董事会所处理的重大决策,每一项都超出了吴恩达一个人能够独自承担的规模。他能够同时兼任这三重角色,原因也在于此。如今的他,已经不再是在一个组织里解决一个问题的人,而是向多个组织传递判断框架的人。

# AGI狂热面前的现实主义

2026年的人工智能行业,正围绕通用人工智能即AGI何时到来这一问题,各自给出不同的数字。埃隆·马斯克设下了十二年的期限,Anthropic的达里奥·阿莫代伊则预测会更早,大约在两三年内就能达到人类水平的智能。谷歌DeepMind的戴密斯·哈萨比斯则给出五年到十年之间的区间。每个人都有各自的依据。模型的性能逐年显著提升,基准测试分数不断突破上一代的极限。然而吴恩达没有站在这一队列之中。他断言,距离真正的AGI还需要几十年的时间。在同僚们纷纷谈论短则几年、长则十年左右的时候,他之所以执意提出不同的数字,原因不在于预测本身,而在于支撑这种预测的措辞用法。

吴恩达质疑的是,AGI这个词已经失去了学术定义。这个词原本指的是与只擅长某一项具体任务的狭义人工智能不同、能够广泛完成人类可胜任的智力工作的系统。然而如今,这个词被平均分配到融资演讲稿、宣传语和政策讨论的修辞之中,本来的含义变得模糊不清。吴恩达用颜色打了个比方来说明这种情况。如果人们把红色也叫蓝色,把黄色也叫蓝色,连绿色都叫蓝色,会发生什么呢。用不了多久,蓝色这个词就什么都指代不了了。哪怕有人说出蓝色,也无从得知那实际上是什么颜色。他的诊断是,AGI这个词正走在同样的路上。随着这个标签被贴到各种水平的系统上,人们一听到AGI这个词所联想到的画面,已经变得因人而异。

这种混乱并不只停留在学术研讨会内部。吴恩达担忧,相当一部分高中生和大学生正因为AGI将很快消灭所有职业这一误解而改变专业。他们是在一项尚未问世的技术的基础上,决定当下的人生方向。企业首席执行官们也不例外。他们依据过度乐观的展望做出资本配置的决定,而当这些决定在几年后被证明错误时,损失将原封不动地落到组织和员工身上。吴恩达真正担忧的地方在此更进一步。他认为,当被夸大的期待无法长久维持、转化为公众的失望时,投资与关注可能会同时抽离。他把这种局面称为人工智能的寒冬。事实上,人工智能研究的历史上曾出现过两次这样的寒冬。期待走在前面,而成果跟不上期待,预算和人才转向其他领域,研究长期陷入停滞。吴恩达认为,如今的这股热潮正冒着重蹈覆辙的风险。

为了廓清言语上的混乱,他提出了一项标准,名为图灵AGI测试。艾伦·图灵此前提出的图灵测试,问的是机器能否通过对话骗过人类。吴恩达修订后的新标准,从另一个起点出发。给人工智能配备一台与人类远程工作者实际使用的完全相同的电脑,也就是浏览器、视频会议软件和文件系统都一应俱全的环境。然后观察它能否在数天之内,像人一样熟练地处理具有经济价值的实际工作。担任评委的人会不断抛出新的任务,试图找出人工智能的极限,即便这样连续多日施加压力,只要没有出现明显逊于人类的地方,才能称之为AGI,这就是他的标准。以这把尺子来衡量,吴恩达的判断是,目前没有任何一个模型跨过了这道门槛。因为在演示视频中像模像样地完成几项任务,与在数日之内不断涌现陌生要求的真实工作环境中稳如泰山地坚持下来,是完全不同层次的事情。

就这样,在拆解围绕AGI这一宏大名号的泡沫的同时,吴恩达并没有背弃当下人工智能已经带来的实质性变化。位于这场变化中心的一个词就是智能体AI。他是较早开始构建和打磨这一概念的人之一。智能体系统所做的事情,不同于人类提出问题、系统一次性给出答案的方式。就像人处理事情要经过多个步骤一样,人工智能也会自行经历若干步骤来完成任务。吴恩达指出,在设计这类系统时,反复出现的框架有四种。回溯并打磨自身产出结果的反思,调用计算器等外部工具的工具使用,把大目标拆解成小步骤并排出顺序的规划,以及不是由单一人工智能、而是由多个人工智能各自处理自己那部分任务后再互相交接的多智能体协作。把这四种框架组合起来,人工智能就能超越简单的问答,去编写并修改代码,审核文件是否符合

关税法规,阅读复杂合同并提炼要点,或者分多个阶段处理客户咨询和医疗支持工作。

问题在于,这个词也开始走上与AGI相似的道路。吴恩达坦言,当他打磨出这一概念并推向世界时,没料到没过多久,各种各样的产品都被贴上了智能体的标签。实际上,连那些不过是按既定顺序调用几个函数的自动化工具,也被贴上了这个名号,于是在宣传语中,这个词的分量再次变轻了。在这一点上,吴恩达反复强调一项原则。实验室里看起来像模像样的演示,和企业每天都要依赖的生产环境,必须用截然不同的标准来判断。研究者给人工智能配上多种工具、让它自由地完成任务时,会出现令人惊叹的场面。然而要把它引入实际公司,情况就不同了。企业要求的是同样的流程运行一万次、十万次,每一次都能得出正确结果的确定性。吴恩达的判断是,当下的人工智能模型在需要自主判断的局面中仍然存在难以预测的部分,毫无管控地放任其运行,信任的鸿沟实在太大。因此在产业现场,人们选择把复杂业务拆解成多个阶段,在每个阶段以稳定的方式实现自动化,再把结果衔接起来。这意味着,在当下这个时间点,分阶段的管控比华丽的自主性更值得信赖。

这里还叠加着另一个悖论。给大语言模型一次性塞进太多工具和说明,反而会出现性能下降的现象。工具列表越长,模型一次要审视的信息量即上下文就越膨胀,注意力从真正需要关注的核心上分散开来,进而调用错误的指令或偏离原本的目标。就连业界广泛使用的模型上下文协议即MCP服务器,也难以摆脱这个问题。如果一次性罗列出几十种工具,模型就会不知该选择其中哪一个,而这种混乱会随着工具的增加而进一步加剧。归根结底,吴恩达指出的结论是,单纯增加工具的数量本身并非上策。他高度评价Anthropic围绕Claude Code打造的开发工具集合即所谓的harness,以及精心构建提示词的技巧,认为它们在实际业务应用中仍然起着决定性的作用。他的观点是,不亚于模型本身性能的,是如何包装这个模型、以怎样的顺序传递指令,这才是决定实务成败的关键。

如此冷静地指出技术局限之后,吴恩达随即把目光转向就业问题。对于人工智能将彻底夺走人类劳动这一担忧,即所谓的职业终结论,不仅是吴恩达,连斯坦福的同事李飞飞等学者也摇头表示不认同。回顾以往的产业革命,新机器与其说消灭了劳动,不如说扩大了劳动的规模和方式,这是他们共同的看法。吴恩达在解决这个问题时,把视线从职业这一宏观单位转移到其中作业这一微观单位上。一份职业实际上是由无数个具体作业汇聚而成的。就拿护士这一份职业来说,其中也混合着照护患者、记录诊疗内容、核对检验结果、与其他医护人员沟通等多种作业。根据吴恩达的分析,在大多数职业中,人工智能能够真正取代的作业份额,大致只在30%到40%之间。反过来说,这也意味着支撑那份职业的60%到70%的领域,依然需要人手。

这个数字所代表的意义不小。回想护士和医生比起直接照护患者这一本职工作,更常被繁重的病历记录之类的文书工作所困的现实,人工智能与其说是抢走这些人饭碗的存在,不如说更接近于把他们从消耗性的重复劳动中解放出来的工具。因此吴恩达提出的核心论断可以这样概括:使用人工智能的人,将取代不使用人工智能的人。这里需要强调的是,吴恩达指出取代的主体并非机器,而是把这台机器熟练地纳入自己工作之中的同事。这意味着,只有在每次出现新工具时都随之调整自己工作方式的人,才能守住下一个位置。这并非头一次发生这样的事情。当电脑和电子表格进入办公室时,同样的分岔路口就已经出现过。有统计显示,未能适应那场变化的人,毕生收入下降,死亡率反而上升,这一事实表明,每逢技术转折期,适应上的差距都可能转化为生活上的差距。

当然,这并不意味着吴恩达认为所有职业都安全无虞。在这一点上,他没有绕弯子,而是给出坦率的答案。他并不回避这样一个事实:确实存在几乎全部业务都可以由人工智能取代的职业类别。他表示,翻译、

配音演员、呼叫中心客服等职业,如今正面临最为明显的威胁。这些职业属于明显偏离前面所说的30%到40%这一平均水平的例外。因为转译语言、出借嗓音、回答固定问题这类工作,正接近于当下的人工智能能够达到与人相当水准的领域。吴恩达并未止步于承认这一事实,而是随即引向社会责任问题。他的立场是,即便某人的工作因技术发展而消失,也绝不意味着这个人可以毫无准备地被赶到街头。他表示,社会有道义上的责任照顾这些人。由此他呼吁政府和教育机构必须切实建立再培训项目和社会安全网。他的意思是,对因技术进步而蒙受损失的人们的关怀,不是可以另行搁置的问题,而应当与技术进步本身当作一对来处理。

越过这场围绕就业的现实讨论之后,吴恩达把目光转向学界围绕人类智能本质展开的争论。杨立昆和戴密斯・哈萨比斯对人类大脑描绘出不同的图景。杨立昆认为,人类大脑是由处理语言的部分、处理视觉的部分、调控运动的部分等高度专门化的多个模块汇聚而成的结果。哈萨比斯则相反,认为人类大脑是一个适应力极强的通用学习器。吴恩达诊断这两种观点之间实际上并不矛盾。他指出,大脑真正令人惊叹之处,恰恰在于可塑性,也就是学习新事物的能力本身。他解释说,一个拿到数学博士学位的人,如今固然在这个领域深度专业化,但如果在不同的环境中以不同的方式接受训练,原本也可能成为国际象棋大师,也可能成为出色的网球选手。这意味着,人类一开始就带着一个通用的学习引擎出生,此后通过所经历的经验,把自己塑造向特定的方向。吴恩达认为,当下的人工智能也正走在与此相似的道路上。一个模型所具有的通用潜力,实际上正在医疗、制造、金融等各自不同的专业领域中,通过解决具体问题而逐渐具备实质内容,成为狭义的应用。

这种乐观的视角,又引向了围绕AGI的另一场争论,即关乎人类存亡的风险争论。埃利泽・尤德科夫斯基之类的人士警告说,AGI一旦被制造出来,整个人类就将陷入危险。吴恩达不认同这种主张。他批评说,尤德科夫斯基的逻辑过于循环,以至于很难判断该从哪里开始反驳。他看待世界的视角,从另一个起点出发。他首先想到的是,如果现在放缓技术发展的步伐,今后仍将有数百万人因癌症、衰老和各种疾病而丧生这一事实。反之,他相信如果大力推进人工智能的发展,就能开辟出拯救更多生命、创造新的财富、把人们从贫困中解救出来的道路。吴恩达的现实主义归根结底同时具备两种态度:一是冷静辨别当下市场炒热的泡沫的眼光,二是坚信这项技术带给人类的利益远大于由此造成的损失的信念。剥去用夸张言辞吹起的期待,与持续推进真正能改善人们生活的工具,对他而言并非彼此不同的方向,而是同一种态度所呈现的两种表情。

再深入看一看各个职业的具体情况,就会发现吴恩达所说的30%这个数字,并非在所有行业中都同等适用。以法律事务为例。在律师处理的工作中,查找判例和挑出合同草案中的错误这类作业,语言模型可以承担相当大的一部分。然而,会见委托人听取情况、在法庭上说服法官和陪审团、在谈判桌上观察对方的表情来协调条件,这些依然是人类的分内之事。制造业现场也呈现出类似的格局。用摄像头挑出不良品的检验作业,视觉人工智能能比人做得更快更准,但布置新设备、留意作业人员的安全、随机应变地应对突发故障,这些是机器难以取代的。吴恩达之所以以作业而非职业作为解决问题的单位,原因就在这里。如果把一个职称之下的多项工作笼统地一并判断,很容易得出草率的结论,但把这些工作逐一拆开来看,人工智能能够深入的位置和人必须坚守的位置,就会分得清楚得多。

这一视角也直接延伸到教育现场。吴恩达表示,大学和职业教育机构需要教给学生的,不是某种特定软件的几种用法,而是判断自己承担的工作中哪一部分该交给机器、哪一部分该由自己负责的眼光。这种判断力,对年长的劳动者和年轻的学生同样必要。如果一名新员工从入职第一年起就把重复性的文书工作交

给人工智能,把那段时间用来经营与客户的关系,那么几年之后,这个人的履历将与那些一直守着文书工作的同事拉开很大差距。吴恩达强调的再培训,在这个意义上,与其说是掌握一门新编程语言的问题,不如说更接近于养成重新规划工作方式本身这一习惯的问题。

智能体工作流带来的变化,也触及公司组织的结构。以往由一个人处理的业务的初始阶段、中间阶段、最后阶段,如今分别由不同的人工智能智能体分工承担,而人则转移到在各阶段之间确认成果、调整方向的角色,这样的案例正在增多。吴恩达认为,这种重组并不只是朝着减少人的岗位这一方向流动。他解释说,反而随着一个人能监督的业务范围扩大,出现了以比以往更少的人手处理更多工作的团队,在这样的团队中,对人的要求也从直接执行实务的人,转变为审核并协调多个智能体工作的人。在这样的组织里,操控人工智能的能力,以及从人工智能给出的结果中挑出错误之处的判断力,正共同成为新的核心能力。

对于这场变化的速度,吴恩达也保持着审慎的态度。硅谷发布会现场展示的演示视频,看起来像是几天内就完成的功能,但那项功能要在实际公司的业务流程中真正扎根,需要长得多的时间。一套新软件通过验证以符合医院或银行的业务规程,并让负责的员工放心使用这个工具,所需的时间,永远比技术本身发展的速度更慢。吴恩达指出,人们常常在这一点上产生错觉。他的意思是,人工智能模型的性能在一年内大幅跃升,并不意味着把这个模型部署到实际业务中的速度也以同样的比例加快。恰恰相反,这个差距越大,技术发展与现场变化之间就越容易出现错位,而急于填补这种错位的尝试,也可能成为进一步吹大吴恩达所担忧的泡沫的原因。

他在达沃斯世界经济论坛之类的场合反复强调的要点,也与此并无二致。他对政策制定者和企业经营者表示,在围绕人工智能的讨论中,必须同时警惕两个极端。一端是认为人工智能几年内就将取代人类一切工作的毫无根据的乐观,另一端是认为人工智能终将把人类推向毁灭的毫无根据的恐惧。他指出,这两种态度都会让人错过当下正在发生的变化,也就是特定作业一项接一项实现自动化、人的角色在其间被重新编排的这场悄然的转移。吴恩达表示,这场悄然的转移,才是当下政府和企业在制定政策与预算时真正需要抓住的对象。他的意思是,比起为尚未到来的超级智能立法预备,如何帮助此刻正在呼叫中心、翻译事务所和工厂里失去工作的人们,才是更为紧迫的课题。

这种态度,也与他反复向学生和实务工作者提出的建议相通。他建议,与其把精力用在预测人工智能在未来几年内会发展到何种程度,不如亲自动手实验,看看用当下手中的工具能做出什么来。因为只有亲身确认过工具局限的人,才能准确指出那个工具力所不及的地方,也就是人依然必须填补的位置。对吴恩达而言,人工智能并非某天终将超越人类的神秘存在,而是此时此刻性能与局限同时显现出来的一件工具。使用这件工具的人的分内之事,在于找出工具做不到的事情,把自己的时间投入到那件事上。这种态度,是支撑他的现实主义的支柱,也是他作为教育者至今坚持的姿态。

接过这个问题,李飞飞从略微不同的角度指出了同样的问题。她在一次采访中说,十年后将只剩下两类劳动者。也就是说,劳动力市场将分化为向人工智能下达指令并审核结果的人,与按照人工智能的指示搬运工作的人。如果说吴恩达先前指出的30%到40%这个数字,是把一份职业内部哪些作业转移给机器细致划分出来的结果,那么这句话更接近于追问这一结果将如何重新分配人与人之间的地位。这意味着,即便在同一家医院、同一家律所、同一座工厂里,熟练驾驭人工智能工具并把握方向的岗位,与确认并补充这些工具处理结果的岗位,也开始明显分化。吴恩达并不把这种区分当作宿命论来接受。他反复强调的再培训必要性,正是源于这样一种信念:在这个分岔路口站在哪一边,取决于训练而非天赋。他判断,要站在下达指令的位置上,就必须有亲身体验过人工智能擅长什么、又遗漏什么的经验,而这种经验的性质,是不分

年龄和学历、任何人都能够积累的。

这场重组,在达沃斯世界经济论坛现场也以另一个名字被讨论过。媒体把它称为人工智能大洗牌,即由人工智能引发的组织大改编。公司的晋升阶梯长久以来都是这样搭建的:新员工先承担重复性实务积累经验,再以这份经验为跳板一级一级往上爬。比如新律师通过审阅数百份合同来培养识别条款陷阱的眼光,新分析师通过手动整理财务报表来练就读懂数字之间关系的感觉。然而,恰恰是这类初级阶段的重复性工作,正是当下的语言模型最先、也最廉价能够取代的领域。吴恩达担忧的地方就在这里出现分岔。他认为,重复性工作一旦消失,新人眼下看似轻松了,但唯有通过那种重复才能培养出来的判断力,培养的机会本身也会随之减少。阶梯的下层一旦交给人工智能,几年后到哪里去培养能爬上阶梯上层的人,就成了新的课题。达沃斯上的讨论把这个问题当作公司必须重新设计新人招聘方式本身的课题来接受。也就是说,不再安排重复性工作,而是从一开始就让新人负责审核并修正人工智能处理的结果,并在这个过程中重新设计训练路径,以培养判断力。

在世界经济论坛安排的场合上,吴恩达没有把这个问题局限在个别企业的人事政策之内,而是将其扩展为整个社会的对话方式。他之所以在这时提出诚实对话这个说法,是因为他诊断当下的公共讨论正被困在两种极端的语言之中。一端是把人工智能说成几年内将拯救或毁灭人类的存在这种语言,另一端则是被这种宏大叙事压制、以至于眼下正在发生的细微变化连新闻都算不上的沉默。吴恩达表示,政策制定者所需要的,是在这两个极端之间某处,如实指出当下哪些作业正在被自动化、又有谁因此实际失去了收入的语言。他表示,只有这种语言站稳脚跟,政府才能及时做出诸如优先把再培训预算分配到哪里、如何调整失业救济标准之类的具体决定。他反复主张,比起为超级智能的到来立法预备,应对呼叫中心和翻译事务所已经在发生的人力调整来编列预算更为优先,这正是从诚实对话这一原则得出的实践性结论。

吴恩达在这场对话中格外用力强调的一点是,围绕人工智能的争论,实际上只在少数几家巨头企业和少数富裕国家内部展开。他依据在AI Fund和Landing AI工作的经验,反复试图证明,无需昂贵的超级计算机和大规模数据,一家制造企业、一家医院也能凭借自己的数据解决问题。他在达沃斯之类的场合向政策制定者提出的要求,也与此并无二致。他认为,讨论的重心应当从能够打造超大型模型的少数几家企业之间的竞争格局,转移到把这些模型拿来、按照自己所在行业的具体问题加以运用的众多企业和劳动者身上。从这个角度看,他关于再培训和社会安全网的呼吁,与其说是提议放慢人工智能发展的速度,不如说更接近于要求不要把这种发展所创造的利益与负担,只集中分配给狭小范围内的人。他相信,既无法阻止技术进步本身,也没有必要阻止,同时又试图让这一进步的果实扩散的速度,与人适应的速度相匹配,这种态度构成了他在达沃斯反复抛出的信息的骨干。

审视这场争论中出现的这些名字,就能窥见为何各人给出的数字互不相同的线索。埃隆・马斯克、达里奥・阿莫代伊、戴密斯・哈萨比斯都是站在最前线、亲自率领团队开发巨型模型的人。他们所处的位置,能够比任何人都更早看到自己实验室里每个月都在刷新的模型性能。因此,他们的预测大多建立在这样一个前提之上:性能曲线今后也将保持与现在相同的斜率继续上升。吴恩达对这种立场所生出的乐观提出了质疑。他反复指出的一点是,实验室里基准测试分数上升的速度,与那种性能真正在医院的诊疗记录系统或银行的授信审查流程中落地的速度,是按照完全不同的时钟运转的。因为基准测试是解答既定题库的考试,而现场却是每天都会冒出新例外的地方。

首席执行官们所犯的资本配置错误,也源于这种时间差。有的公司相信几年内就能把全部咨询人力转移给人工智能,于是先削减招聘。可等到实际引入系统后才发现,十次里九次能顺畅给出答案,却仍有一次会

出现给出离谱指引的事故。这种错误率在演示舞台上并不显眼,但在每天要处理数万件咨询的呼叫中心里,却每天都在积累为反复出现的投诉。最终,公司要么不得不重新补回被削减的人力,要么被迫在事后重新设计人机协作的流程。吴恩达所说的资本配置的失误,正是源于这种把性能与信任之间的差距看得过窄而做出的决定。

图灵AGI测试之所以特意要求持续数天的考验,也与这种信任的鸿沟相关。对一个问题给出一个像模像样的答案,与撑过从周一到周五连续不断的实际工作周,两者要求的东西并不相同。周二处理事情的方式,能否原样适用于周三提出的略有不同的要求,会议日程突然改变时能否察觉这一变化并自行调整后续作业,评委故意抛出模糊指示时能否反问确认,诸如此类的问题每天都会被重新抛出。要通过这样的考验,需要的不是一次出色的回答,而是数日之内不动摇的一贯性。吴恩达断言当下的模型中没有一个跨过这道门槛,其依据正在于这种一贯性的缺失。

他对智能体工作流所持的审慎态度,也源自同一根源。反思、工具使用、规划、多智能体协作这四种框架,各自都开启了强大的能力,但这些框架叠加得越多,系统在中途岔到错误方向去的通道也随之增多。规划阶段一旦把目标定错,后续所有工具调用从一开始就会全部落空,而在多个智能体相互协作的结构中,也会出现一个智能体留下的错误判断被下一个智能体原样继承、进而放大错误的情况。因此,吴恩达曾经共事过的产业现场团队,不是把这四种框架一次性自由地放开,而是选择在每个阶段都留下人可以确认的节点。也就是稍微降低一点自由度,来守住一旦出错就能准确找出问题所在的结构。

工具越多性能反而下降的这一悖论,也与这种设计原则相通。就像一个人到新单位上班的第一天,如果一次性收到几十本规章手册,反而容易漏掉真正重要的规定一样,语言模型一次要处理的工具列表一旦变长,也更难从中挑出此刻真正需要的那一个。因此,在实务中取得成效的团队,不是一次性把所有工具都交给模型,而是从中挑出当下所承担业务真正必需的几种工具再交给它。吴恩达之所以关注Anthropic围绕Claude Code搭建起来的这套harness,原因也在这里。这套harness不是一次性把一切都交给模型,而是具备在当前这个阶段只挑出所需信息和工具交出去、再把结果传递到下一阶段的结构。他反复确认的一点是,不亚于模型本身固有性能的,正是这种筛选并传递信息的设计,决定着实务的成败。

在围绕就业的讨论中,也反复出现类似的原理。30%到40%这个数字,大体上与一份职业内部定型化业务,也就是按既定流程反复进行的业务所占的比重相重合。整理文书、转录记录、对固定问题给出固定答案这类工作,恰好是语言模型擅长处理的形态,而在陌生情境中随机应变地判断、通过读懂人的表情和语气来建立信任,则依然是人的分内之事。翻译、配音演员、呼叫中心客服之所以尤其面临高风险,是因为这些职业所承担的大部分工作,都接近于把既定语言转换成另一种语言、把既定台词转成声音、把既定问题转成既定答案这类定型化作业。吴恩达之所以单独点出这三类职业并直呼其名,是为了不掩盖这样一个事实:适用于其余职业的30%这一平均值,并不适用于它们。

围绕大脑可塑性的说法,也与这种定型化与非定型化的区分相互呼应。吴恩达说,一位数学博士如果接受了不同的训练,原本也可能成为国际象棋大师或网球选手,他这样说是为了说明人类的智能从一开始就不是固定在某一种狭隘用途上的。当下的人工智能模型最初也是以处理语言的通用能力起步,但在被放置到医疗、制造、金融各自的现场后,便经历了按照那个行业的语言和流程被打磨的过程。这个过程,与一个人大学毕业后在特定单位度过几年、逐渐熟悉那个行业的方式相似。吴恩达之所以把当下的人工智能说成是一段熟悉的学习过程,而非一次神秘的飞跃,背后正是他长期以来对这种相似性的观察。

尤德科夫斯基提出的存亡争论,也可以在这一脉络中重新解读。吴恩达之所以把那种主张称为循环论证,是因为那种逻辑先已经把人工智能是危险的这一结论作为前提摆在那里,之后再把支撑这一前提的依据事后补上去。相反,吴恩达自己的判断,是从眼前的数字出发的。他先数一数每年因癌症和衰老而离世的人数、技术每延迟一分就会增加的那个数字,然后再把人工智能降低这个数字的可能性,与制造新风险的可能性并列权衡。在这番权衡中,他把砝码放在利益大于风险这一边。他在达沃斯要求进行诚实对话,归根结底也是要求把这番权衡,建立在此刻可以计数的数字和此刻正在发生的变化之上,而不是抽象的恐惧或含糊的救赎式语言。

# Support This AI Library Book

If this book was useful, you can support future translations, public editions, and open AI knowledge projects through PayPal.

**PayPal**

**https://paypal.me/kimkjj**

Scan the QR code or open the link above.