



AI领域2017进展总结

AI前线特刊



卷首语

人工智能的未来已来

策划 | Tina

撰稿 | Vincent

说到近几年工智能发展的开端，我们又要说起那个已经被翻烂了的起源话题：李世石大战 AlphaGo，这是近三年以来，人工智能头一次以如此耀眼的方式成为了各大媒体的头条，继互联网之后，人工智能仿佛洪水一般淹没了整个世界——智能时代到来了。

2017 年，大数据掀起的热潮开始逐步让位于人工智能——事实上，AI 已然成为本轮技术漩涡的核心所在。

技术上来说，2017 年始于 Pytorch，并对 Tensorflow 构成真正意义上的挑战，特别是在研究方面。对此，Tensorflow 通过在 Tensorflow Fold 中发布动态网络迅速作出回应。大玩家之间的“AI 之战”轰轰烈烈，其中最激烈的战争均围绕云而展开，所有的主要供应商都已经在各自的云服务中加紧布局 AI。亚马逊已经在他们的 AWS 进行大量创新，比如其最近推出构建和部署 ML 模型的 Sagemaker。另外值得一提的是，小型玩家也在不断涌入，例如 Nvidia 最近推出了他们的 GPU 云，位训练深度学



习模型提供了另一个有趣的选择。虽然战况激烈，但我很高兴看到各行业在必要时能够凝聚在一起。另外，新的 ONNX 神经网络表达标准化是实现互操作性重要且必要的一步。

从应用场景上看，这一年，AlphaGo 再一次击败了人类的冠军棋手，之后又通过自我训练成功进化出 Alpha Zero；自动驾驶成为热潮，不论是特斯拉、Uber、百度、谷歌这样的大厂，还是图森、景驰等等这样的初创企业，纷纷投身自动驾驶；此外，人工智能芯片成为了新的“战场”，前有英伟达、谷歌，后有地平线、寒武纪，纷纷在 2017 年即将结束之时放出大招，吸睛又吸金……

除了上述案例，人工智能在 2017 年还有太多值得铭记的瞬间，在此暂不一一赘述。从以上案例中，我们可以很明显的感受到：AI 技术的核心正在从学术界逐渐转向工业界。到 2017 年年底的时候，工业界已经成为了推动 AI 发展的最主要力量。除了原本就以技术为核心的企业外，还有一些看似和 AI 完全搭不上关系的企业也开始在业务中加速布局 AI，可

见这一轮人工智能浪潮的影响已经扩散到了各个领域，并还将持续下去。

进入 2018，更多的机会在等待着 AI 从业者。有专家预测，2018 年必将成为全球人工智能优先发展的突破之年。随着中国和加拿大的人工智能和印度等国家从 IT 向人工智能转变，来自欧洲、亚洲、印度、沙特阿拉伯等国家，乃至全球对 AI 的需求将继续增长。美国和海外对企业培训的需求都很大，人工智能将实现大规模提高效率，传统行业如制造业、医疗保健和金融将会从中受益。人工智能创业公司将向市场推出新产品，并全面提高投资回报率。而机器人、自动驾驶汽车等新技术将会带来惊人的进步。

虽然未来充满希望，然而一些现有的问题却不容乐观：部分 AI 技术应用落地和普及仍然困难、全球 AI 和大数据人才依旧短缺、各国法律政策对 AI 还是没有十分明确的规定……

这注定是一个挑战与机会并存的时代。2017 年已经成为过去，在 AI 领域又太多里程碑值得纪念，总结 2017 是为了更好的迈向 2018，所以 AI 前线在 2018 年之初为各位读者奉上这样一本迷你书，涵盖了来自全球 AI 和大数据领域技术专家的年终总结与趋势解读，同时还有世界知名技术大厂的年终技术总结与趋势预测。

如果你是 AI 从业者，希望这本迷你书能够为你的职业生涯规划做出一定的贡献；如果你不是 AI 从业人员，那么这本书是你了解 AI 发展趋势的绝佳参考，一定不要错过。

聚焦最新技术热点 沉淀最优实践经验

[北京站]2018

北京·国际会议中心

演讲：2018年4月20–22日 培训：2018年4月18–19日

精彩案例 先睹为快

《Netflix的工程文化：是什么在激励着我们？》

Speaker: Katharina Probst

Netflix 工程总监

《Apache Kafka的过去，现在，和未来》

Speaker: Jun Rao

Confluent 联合创始人

《人工智能系统中的安全风险》

Speaker: 李康

360网络安全北美研究院负责人, IoT安全研究院院长

《从C#看开放对编程语言发展的影响》

Speaker: Mads Torgersen

微软 C#编程语言Program Manager

《Lavas：PWA的探索与最佳实践》

Speaker: 彭星

百度 资深前端工程师

《浅谈前端交互的基础设施的建设》

Speaker: 程劭非（寒冬）

淘宝 高级技术专家

《深入Apache Spark流计算引擎：Structured Streaming》

Speaker: 朱诗雄

Databricks软件开发工程师, Apache Spark PMC和Committer

《AI大数据时代电商攻防：AI对抗AI》

Speaker: 苏志刚

京东安全 硅谷研究中心负责人

《QUIC在手机微博中的应用实践》

Speaker: 聂永

新浪微博 技术专家

《阿基米德微服务及治理平台》

Speaker: 张晋军

京东 基础架构部服务治理组负责人, 架构师

8折 优惠报名中，立减**1360元**
团 购 享 受 更 多 优 惠

访问官网获取更多前沿技术趋势

2018.qconbeijing.com

如有任何问题，欢迎咨询

电话：15110019061，微信：qcon-0410



ArchSummit

全球架构师峰会

2018 · 深圳站

从2012年开始算起，InfoQ已经举办了9场
ArchSummit全球架构师峰会，有来自Microsoft、
Google、Facebook、Twitter、LinkedIn、阿里巴
巴、腾讯、百度等技术专家分享过他们的实践经验，
至今累计已经为中国技术人奉上了近千场精彩演讲。

- **2017.07.07-08 深圳站**
how to use sagas to maintain data consistency
in a microservice architecture
---Chris Richardson, *POJOs in Action* 作者，知名微服务专家
- **2017.12.08-11 北京站**
《创新是人类的自信》
---王坚博士，阿里巴巴集团技术委员会主席
- **2018.7.06-09 深圳站**
限时**7折报名中**，名额有限，快快抢购。

7折报名中
名额有限，快快抢购

华南地区架构领域最有影响力的会议，届时有哪些专题和演讲，敬请扫描右方二维码浏览官网。



目录

专家总结与趋势预测

8 数据科学与机器学习：2017 年主要成果与 2018 年核心趋势

18 不止 Google vs. Nvidia：深度学习引领 AI 芯片大战

28 人工智能落地，商业化比技术更重要

AI 名人名企回顾

42 DeepMind 的 2017 年：历程回顾

47 阿里巴巴年度技术总结：人工智能在搜索的应用和实践

56 年终解读：2017 年的语音识别，路只走了一半

64 2017 年，吴恩达都干了些啥？我们替他做了份年终报告

AI 开发者

72 2017 开发者盘点：在他们眼中，AI 是吹牛灌水还是真的来了？



数据科学与机器学习： 2017 年主要成果与 2018 年核心趋势

作者 | Kirk D. Borne

译者 | 核子可乐译

展望 2018年，我们应该会迎来真正的 AI技术发展动力，而不仅仅是相关炒作。现在是时候验证 AI的实践人地，衡量其投资回报率并真正实现其可行性了。在新的一年中，AI的发展方向与 2017年的重点领域不会有太大差别，仍主要集中在流程自动化、机器智能、客户服务、深度个性化以及劳动力转型层面。另外，物联网技术也将日益成熟，包括带来更理想的安全功能、模块化平台、用于访问传感器数据流的 API以及边缘分析接口。另外，我们还可能看到这对数字化双雄在制造业、公共事业、工程及建筑行业当中变得更加主流。我同样相信，到 2018年，更多从业者将需要肩负起把 AI的积极效益传达给抱有怀疑态度的公众这一责任。

Tom Davenport, 巴布森学院信息技术与管理学杰出教授, 国际分析学院联合创始人, 麻省理工学院数字化经济学倡议研究员, 德勤公司高级分析顾问。

2017 年的主要进展

企业 AI 走向主流: 众多大型企业都已经开始推进自己的人工智能或机器学习计划。一部分企业甚至已经拥有超过 50 个采用各类技术的研发项目, 但其中大多数属于“粗放性”项目, 即发展目标相对有限。2017 年出现了一种新的倾向, 即越来越多大型供应商开始将其“变革性”产品转向开源, 允许参与者以 DIY 方式推动项目发展。当然, 这也意味着企业用户必须启用或培养出高水平数据科学人才, 方可与这样的背景性趋势接轨。

机器学习被应用于数据整合: 数据管理与分析领域中这一历史最为悠久的难题正在机器学习的助力下得到解决。原有劳动密集型数据整合与管理方法正在被目前各类数据库所采取的, 类似于对数据元素进行“概率匹配”的机制所取代——或者至少加以补充。利用此类工具(通常由外部工作流以及众包技术专家负责支持)有望将数据整合时耗降低至原本的十分之一。

保守企业接受开源: 银行、保险及医疗卫生等行业当中的传统保守企业, 如今也开始积极采用开源分析、人工智能与数据管理软件。其中一部分企业主动劝阻员工继续使用专有型工具; 也有部分企业允许员工自由作出选择。这种对开源技术的接纳态度, 一方面源自成本构成因素的转变, 另一方面则源自开源项目出色的实际效果以及高校毕业生们的选择取向。

2018 年发展预期

我们已经进入“后算法”时代: 从历史角度看, 分析师与数据科学家们需要掌握大量相关知识, 从而判断哪种算法能够达成既定目标。但时至今日, 分析与机器学习流程自动化已经能够在处理过程中引入成百上千种不同算法, 而且更重要的是考量某一模型或者模型集合的实际表现。尽管目

前还没有实际发生，这最终很可能带来“民间数据科学家”的兴起。

独立 AI 初创企业的吸引力开始减弱：在风险投资基金的推动之下，过去几年中已经先后涌现数百家 AI 初创企业。这些公司往往仅能够解决相对有限的问题。然而，即使其工作成果富有成效，但大多数企业客户仍然很难将其方案同现有流程及系统整合起来。因此，成熟的企业更倾向于自主开发相对易于整合的 AI “微服务”，或者直接从供应商处采购内嵌有 AI 内容的系统方案。

Jill Dyche, @jilldyche, @SASBestPractice 公司副总裁。商业畅销书作者。

时至今日，每一家公司都在涉足人工智能或者机器学习领域。纵观 2017 年，事实证明是金子总会发光——而各类供应商，包括很多我认为跟 AI 根本搭不上边的企业，都希望借此让自己的产品组合更上一层楼。而颇具讽刺意味的是，这些产品确实非常新颖，因此其中不少供应商很可能彻底超越自身原本的业务定位。

2018 年的 AI/机器学习用例与商业对话还将继续增长。为什么？理由很简单，大多数企业管理者属于需要解决实际问题的商人，他们并不关心神经网络与稀疏数据要如何接驳在一起。他们对自然语言处理中的词汇推理挑战也不抱任何兴趣。相反，他们只是希望加速自身供应链运作速度，了解客户要做什么 / 买什么 / 说什么，并由计算机给出答案。这显然属于规范分析的范畴，而有能力以更低门槛交付相关解决方案的供应商将在新一年中占据绝对优势。

Bob E. Hayes @bobehayes, 研究员兼作家, Business on Broadway 发行人，拥有工业组织心理学博士学位。

数据科学与机器学习功能正越来越多被各类行业所采用，并渗透至多种应用领域。

2017 年，我们见证了 AI 功能的巨大进步。尽管目前的现有深度学习模型仍然需要大量数据进行算法教授，但在神经网络与强化学习的双重支持之下，事实证明我们已经能够在无需依赖数据集的前提下构建起高性能算

法。DeepMind利用这些技术并创建起 Alpha Go Zero——该算法超越原有成果的优势在于，其只需要自我对弈即可快速提升棋力。

随着人工智能技术在刑事司法、金融、教育以及办公环境下的持续普及，我们将需要建立算法标准以评估由其带来的不准确性与偏见问题。另外，对于 AI可能带来的社会影响的关注亦在持续升温，其中包括建立规则以约束 AI的使用条件（例如避免出现决策‘黑匣子’）以及理解深度学习算法如何作出决策。

即使对于诞生于互联网时代下的企业（例如 imgur与优步等），安全违规问题同样不容忽视。因此，我们将看到技术行业在彻底扭转安全实现方法层面所作出的努力——具体包括提高区块链（虚拟分类账）可见性以作为改善企业保护其管辖范畴内各类数据的可行方法。

Carla Gentry, Analytical Solution公司数据科学家, @dara_nerd。

2017年，每个人都开始讨论机器学习、人工智能与预测分析。遗憾的是，很多企业 /供应商只是将其作为新的“流行语”，或者说将其作为自我标榜的手段——而非真正认真积累专业知识。如果参阅 Twitter上与此相关的“热门话题”，大家肯定会感到困惑，因为又是同一批人与上年一样在媒体上发布大量营销帖子！要在这些领域有所建树，我们真正需要的经验源自时间与人才方面的投入，而不仅仅是“行为呼吁”和炒作……一如既往，经验仍然极为重要！

在我看来，2018年将成为数据科学与预测分析领域的真正起效的一年。这不仅是因为趋势所致，更是因为这些技术将能够切实让我们的业务发生巨大变化。预测性招聘能够帮助企业节约数百万美元的人才流失与周转损失；AI与机器学习能够在几秒钟内完成原本需要大量时间投入的工作！然而，虽然技术能够让我们带至新的高度，但我们也要牢记自己身为人类的本分。作为数据科学家或者技术作者，我们应当在人性与道德实践层面以不偏不倚的态度践行开诚布公的指导原则。

Gregory Piatetsky-Shapiro, KDnuggets公司总裁，数据科学家，KDD大会与知识发现与数据挖掘专业组织 SIGKDD联合创始人。

2017 年的主要进展

2017年内，AlphaGo Zero成为 AI领域最重要的研究进展。

数据科学自动化水平持续提升，更多工具的涌现令机器学习平台实现自动化升级。

AI炒作与预期的升温速度高于 AI与深度学习的成功速度。

2018 年关键性趋势展望

GDPR（欧洲通用数据保护条例）将于 2018年 5月 25日生效，这将对数据科学产生重大影响，其中对于解释权（您的深度学习方法能否解释为何拒绝某人的贷款要求？）以及偏见与歧视预防提出要求。

谷歌 DeepMind团队将继续推进 AlphaGo Zero的发展，旨在使其在另一项几年前被普遍认为无法由人工智能完成的任务当中取得同样惊人的表现。

我们将看到自动驾驶车辆（包括轿车与卡车）的持续发展，包括解决各类紧急问题（例如在拉斯维加斯试运行的自动驾驶车辆不会让路等）。

AI泡沫将继续存在，但动荡与巩固的迹象亦已经出现。

Dr. GP (Ganapathi) Pulipaka, @gp_pulipaka, DeepSingularity LLC
首席执行官兼首席数据科学家。

2017 年年内的机器学习、深度学习与数据科学发展

AlphaGo Zero代表着一种新的强化学习实现形式，其能够自我指导，且无需人工干预或者历史数据集的参与。

Python（共拥有 165万次 GitHub推送）、Java（共拥有 232万次 GitHub推送）以及 R（16万 3807次 GitHub推送）已经成为 2017年年内最受欢迎的编程语言。

在 CPU上执行神经网络功能时所需要处理的规模化大数据可能随着时间推移带来极为可观的功耗成本。谷歌公司发布了第二代 TPU，其中采

用的精确设计工程成果包括将协处理器接入通用 PCIe 决线、利用乘法累加器（简称 MAC）处理流量，从而复用寄存器中的值进行数学计算，最终借此节约价值数十亿美元的能源使用成本。

英伟达公司推出基于 Volta 架构的 Tesla GPU，用于以每 GPU 高达 120 万亿次的峰值计算性能助力深度学习与机器学习负载。

关于量子计算的炒作关注点，已经由 D-Wave 转向采用 Python 语言的 QISKit 量子编程堆栈的 20 量子位计算机。

2018 年机器学习、深度学习与数据科学发展趋势

McAfee 实验室发布的 2018 年威胁研究报告显示，对抗性机器学习技术将在无服务器环境下为网络入侵检测、欺诈检测、垃圾邮件检测等网络安全领域内的重要事务带来强大的机器处理速度加持。

HPE 公司将开发点积引擎，同时面向深度神经网络、卷积神经网络以及递归神经网络推出自己的高性能神经网络计算芯片。

量子机器学习的未来命运取决于能够将量子位提升至 10 以上，同时在同一量子位内容纳多达 100 种维度。未来将会出现大量利用量子位技术制造的微型芯片，而这将最终成就令人叹为观止的量子计算机。

2018 年，物联网与边缘计算同机器学习相结合所面临的障碍将得到进一步缓解。地理空间智能方案将在移动手机、RFID 传感器、UAV、无人机以及卫星等所承载的突破性算法的支持下得以实现。

自监督学习与自主学习将为机器人提供更为新颖的深度学习技术助力，这意味着机器人将能够与周边地面乃至水下环境进行顺利交互。

Paul Gearan, Heather Allen 与 Karl Rexer，领先数据挖掘与高级分析咨询企业 Rexter Analytics 公司骨干。

时至今日，我们仍然很难确保不具备研究或者分析专业知识背景的用户群体随时随地有效运用商业智能软件。虽然存在 Tableau、IBM 沃森以及微软 Power BI 等软件方案，但根据 Rexter Analytics 公司于 2017 年收集的数据显示，只有略高于半数的受访者表示数据科学团队以外的人员能够使

用此类自助服务工具。而在使用此类工具时，约有 60% 的受访者表示难度很大，且最常见的问题集中在无法把握分析过程以及无法准确理解分析结果方面。

2018 年，实现“数据科学工具民主化”这一承诺对于扩大分析的普及度与积极效用，最终带来有效且有意义的结果而言至关重要。一般来讲（根据我们的经验），我们仍然需要一支具备综合性多学科专业背景的团队，他们将帮助那些不具备分析类培训经验的员工及管理层利用相关工具对其假设结论进行探索与可视化处理。但同样重要的是，该团队还需要开发模型并与数据科学专业人士一同进行结果解释，其中数据科学专家将凭借培训经历以准确判断特定分析技术的适用范围与局限性。

Eric Siegel，预测分析全球大会创始人。

2017 年，机器学习迎来了全速推进的三大发展趋势，相信这一点将在 2018 年继续保持下去。在这三种趋势中，有两种值得肯定，但有一种则为“必要之恶”：

机器学习在各类领域中开始以商业应用形式快速普及 -- 具体包括市场营销、财务风险管理、欺诈检测、劳动力优化、制造业以及医疗卫生等等。感兴趣的朋友不妨关注将于 2018 年 6 月于拉斯维加斯召开的预测分析全球大会，借此了解这一广泛性发展趋势以及已经有哪些领先企业借此实现实际价值。

深度学习在流行充与实际价值两个方面大放异彩。这种相对年轻的先进神经网络方法将机器学习扩展到了新的水平——即在大信号输入类问题方面提供高成效，具体包括图像分类（自动驾驶车辆与医学图像解析）、声音（语音识别与发言人识别）、文本（文档分类）甚至是高维度点击流处理等“标准”业务问题。

遗憾的是，人工智能仍存在过度炒作问题。尽管专业人员有时也会使用 AI 这一定义尚不明确的术语指代机器学习，但各类分析服务供应商以及媒体则会刻意沿用这一称谓以暗示某些明显不切实际的能力，甚至误导受众产生远超真实状态的期望。正如 Arthur C. Clarke 所指出，“任何足够

先进的技术都像是种魔法”，但这并不代表着我们能够想到或者曾经出现在科幻作品中的一切“魔法”都有办法通过技术手段来实现。AI最终将实现自我意识，甚至可能出于武断甚至恶意立场给人类生存构成威胁等结论实际上只是一种鬼故事——众多厂商之所以对此大加宣扬，无非是为了对机器进行拟人化甚至是神化，从而推动产品销售。这里建议大家对“AI”概念抱以更为轻松的心态，而不应偏听偏信一些纯属子虚乌有的胡说八道。

Jeff Ullman，斯坦福大学计算机科学系荣誉教授。他关注方向包括数据库理论、数据库集成、数据挖掘以及基于信息基础设施的教育工作。

我最近在一次会议上与 John Hopcroft以及 Al Aho两位老同事再度聚首。我的发言并没有新鲜之处，但 Al与 John却谈到了不少人们普遍关心的话题。

John (Hopcroft)谈到了深度学习算法的分析机制。他曾经进行了一些实验，希望以不同顺序利用同一组数据进行训练以观察对网络节点活动产生的影响。他最终发现一部分网络节点会在不同训练顺序之下始终保持同样的行为。但在另一些情况下，虽然无法实现节点到节点映射，但单一网络中的某些小型节点集会与另一网络中的另一节点集表现出同样的影响。这项工作尚处于初始阶段，但我大胆就此作出一点推测：

对深度学习网络分析机制的深入探究将帮助人们真正理解深度学习技术，包括其使用方法与潜在陷阱。

接下来，Al Aho则谈到了量子计算。目前以 IBM、微软、谷歌等为代表的众多全球巨头级企业都在投入大量资金构建量子计算机。此类设备拥有多种不同实现方法，但令 Al感到兴奋的是他的一名学生已经在微软公司打造出可观的成果——构建一套编译器与模拟器，用以设计量子算法并完成测试。这意味着即使不存在真正的量子计算机，现在我们也已经能够立足模拟器体验量子计算的强大能力。这不禁令我想起上世纪八十年代的集成电路设计工作。当时我们同样使用编译器——负责将高级语言转化为能够由电路所理解的机器码，而后进行模拟。这样作的好处是大家可以尝试不同的算法，而无需耗费大量时间构建物理电路结构。值得强调的是，

在量子计算层面，构建物理结构的问题已经不只是“缓慢而昂贵”，而是“根本不可能”。实际上，我怀疑量子计算在短时间内还无法成为现实，但为其投入资金并设计相关算法仍然很有必要。正如 AI 所指出，在过去一年中，效率更高的量子计算线性代数算法取得了令人振奋的进展。如果这一目标最终实现，相信数据科学家们将对量子计算抱有更为坚定的信心。而这也引出了我的下一项预测：

即使真正的量子计算机在短时间内——甚至需要数十年——始终无法实现，量子计算乃至与数据科学相关的算法，仍将在未来几年内得到更多关注。

下面我再加入我个人作出的一项更接地气的预测！

由 Hadoop 转向 Spark 的趋势将持续下去，并最终令 Hadoop 消失在人们的视野当中。

Jen Underwood, Impact Analytix LLC 创始人，公认的分析技术专家，在产品管理与设计方面拥有独道见解，且具备 20 多年的数据仓库、报告、可视化与高级分析解决方案实践开发经验。

回顾 2017 年，我会将其视为智能分析平台全面崛起的一年。从分析机器人到自动化机器学习，数据科学中的各个层面都开始涌现出大量极为复杂的智能自动化功能。数据整合与数据筹备平台如今已经拥有出色的智能化水平，能够以即插即用方式对接数据源，在数据管道发生错误时实现自我修复，甚至可利用从人类交互中学习到的知识自主进行数据质量维护或管理任务。增强型分析产品已经开始初步实现机器学习技术的民主化承诺。最终，能够率先提供预打包最佳实践算法设计蓝图以及一定程度自动化特征工程能力的平台将成为数字化时代下分析工具领域的游戏规则改变者。

面向新的一年，我希望自动化人工智能方案能够以无缝化方式被统一至更多分析与决策流程当中。随着企业对此类方案的采用，我想可能会有越来越多的人对自动化决策模糊的实现方式以及如何确保其以负责任方式对真实世界中的问题提供指导而感到担忧。欧盟通用数据保护条例所规定

的最后期限要求我们抓紧时间开放这些分析层面的“黑匣子”，确保其得到正确使用，最终以可靠且尽职的方式对个人数据加以管理。

作者介绍

Kirk D. Borne, @KirkDBorne, 博思艾伦公司首席数据科学家, 天体物理学博士, 顶尖数据科学与大数据意见领袖。



唐杉：2017，AI 芯片元年

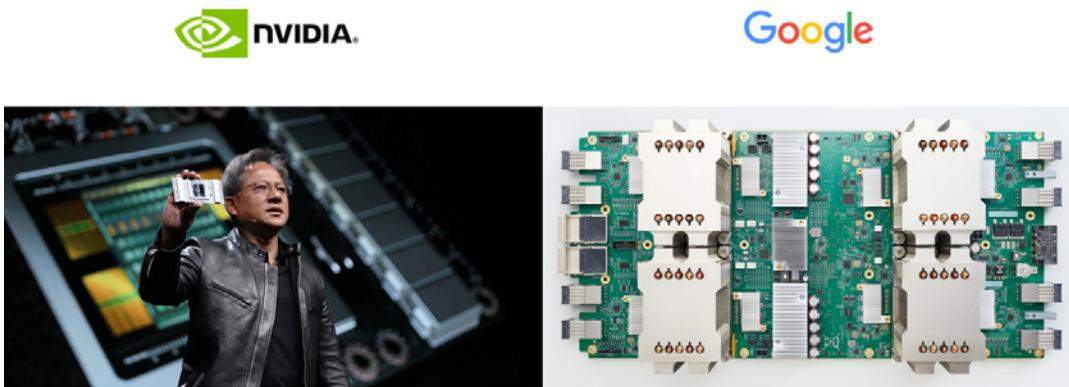
作者 | 唐杉

2017 年，AI 芯片是半导体产业的亮点，而它受到的关注又远远超出半导体的圈子。这一年，从科技巨头到初创公司，新老角色轮番登场，为我们上演了精彩好戏。若干年后，当我们再回头来看，一定可以把 2017 年作为 AI 芯片元年。

Goole vs Nvidia · 巨头之间的“错位战争”

四月初，Google 公布了一篇即将在 ISCA2017 上发表的论文：“In-DatcenterPerformance Analysis of a Tensor Processing Unit”。可以说正是这件“小事”，揭开了一部年度大戏的序幕，而它产生的深远影响甚至可能会持续到很多年之后。其实，在 2016 年 6 月的时候 Google 就透露了自己研发了一款在云端使用的专用 AI 芯片，TPU (Tensor Processing Unit)。Google 做 AI 芯片当然是吸引眼球的新闻，但苦于一直没有公布细节，大家也只能猜测和等待。因此，这篇普通的学术论文，得到了媒体的极大关

注。我也在第一时间写了一篇评论文章：“Google TPU 揭秘”，也是我的公众号阅读量最大的文章之一。对 TPU 高度关注的当然不只我们这些吃瓜群众，还有 AI 芯片领域绝对的统治者 Nvidia。后面就发生了黄教主和 Google 间关于 TPU 的 Benchmark 结果是否合理的口水战。而早在 2016 年 Google 透露 TPU 的时候，Nvidia 就多次表示它对 GPU 在 AI 运算上的统治地位没有什么威胁。



5月11日，Nvidia GTC2017大会，黄教主在Keynote上抛出了最新的GPU Volta (GV100)。Nvidia股票应声大涨，媒体也是大肆报道。AI芯片的焦点似乎又回到了Nvidia这一边。

除了公布了重量级的Volta，GTC上还有一个“小事件”，Nvidia宣布开源它的DeepLearning Accelerator (DLA)，9月正式公开。这个发布，在黄教主的Keynote中是一句话带过，但在业界引起的震动却一点也不小。“Nvidia为什么要搞开源？会开源什么东西？这个开源会不会影响众多初创公司的前景？”对这些问题的讨论一直延续到NVDLA真正开源之后。

没过多久，5月17日，在Google I/O大会上，Google公布了第二代TPU，用媒体的话说“…stole Nvidia’s recent Volta GPU thunder…”。虽然TPU2的细节公布的并不多，但指标确实看起来很不错，而且具有非常好的可扩展性。唯一的遗憾就是它并不对外销售，只能以TPU Cloud的方式

供大家使用。

9 月下旬，Jeff Dean 这位 Google 的软件大神参加了 HotChip 这个芯片界的重要会议，并在 Keynote “Recent Advances in Artificial Intelligence via Machine Learning and the Implications for Computer System Design”也亲自介绍了 TPU 和 TPU2 的情况，把它们作为新的计算生态中重要的一环。

9 月底，NVDLA 在承诺的最后期限之前开源了 NVDLA 的部分硬件代码，同时公布了未来开源更多硬件和软件的路线图。这之后，大家对 NVDLA 也做了各种分析和讨论，试图把它玩起来。从目前来看，NVDLA 的开源好像并没有影响众多初创公司的融资。这个话题我们后面再说。至于 Nvidia 开源 DLA 的原因，官方的说法是让更多人可以更容易的实现 Inference，促进 AI 的推广，特别是在众多嵌入式设备上的应用。但从整个开源的过程来看，这个开源的决定似乎是比较仓促的。DLA 来自 Nvidia 自动驾驶 SoC 中的一个 module，最初并不是以开源 IP 为目的而设计的。而且 9 月的开源也只公开了一部分硬件代码和相应的验证环境，离真正能用起来也还是有较大差距。我们不好判断这个开源的决定是否和 Google TPU（在 Inference 上有比较大的优势）的强势亮相有关系。但基本的推测是，在 Deep Learning 中 Nvidia 的核心利益应该在于 Training（目前 GPU 还是 training 的最好平台）。让 Inference 门槛更低，渗透到更多应用，特别是 Edge 端，从而进一步促进 Training 的需求，应该是符合它的最大利益的。而且 NVDLA 的软件环境还是使用 Nvidia 的 CUDA/TensorRT，还是由 Nvidia 掌控的。

这场从一篇论文开始，几乎贯穿了 2017 年全年的 Google 和 Nvidia 的明争暗斗，对业界的影响可能要远远超过这两家公司本身。我之所以把它称为“错位”的战争，是因为它发生在 Google 这样的传统的软件巨头和 Nvidia 这样的芯片巨头之间。如果换成 Intel vs Nvidia，似乎是再正常不过的。Google 的参战，也许是开启了新的时代。我们可以看到，不仅是 TPU，Google 在 10 月又公布了他们在“Google Pixel 2”手机中使用的定制

SoC IPU (Image Processing Unit)。和 Apple 越来越多的自己定制芯片一样，Google 这样的科技巨头同样有应用（明确知道自己要什么），技术（对相关技术的多年积累），资源（不缺钱，不缺人）上的优势，定制自己的硬件，甚至芯片会变得常态化。同时我们也看到，Google TPU 的示范效应已经显现，更多的科技巨头加入 AI 加速硬件的竞争。Tesla 宣布自己定制自动驾驶芯片；Amazon，Microsoft，以及国内的 BAT，华为都在 Cloud 中提供专门的 FPGA 加速的支持；据称 Big Five 中还有在自己开发芯片的；BAT 也都在组建芯片设计的团队，等等。虽然大家具体的架构和实现方式不同，但都反映出对 AI 专用硬件的极大兴趣。相信未来这一趋势会越来越明显。

同时，传统的芯片巨头当然不会坐视这个巨大的市场被 Nvidia 主宰或者被 Google 们瓜分。Intel 连续收购了 Nervana（云），Movidius（端），Mobileye（自动驾驶），Altera（FPGA），又把 AMD 的 RajaKudori（GPU）招至帐下，甚至还搞了 Loihi（neuromorphic），可以说拿了一手好牌；虽然动作没有大家想象的那么快，但后面的发力还是值得期待的。AMD 也在努力追赶，毕竟他们的 CPU+GPU 有自己绝活，而整个公司也已经逐渐走出了低谷。而且，不管 Tesla 和 AMD 合作自动驾驶芯片的消息到底是真是假，芯片公司这种输出芯片设计能力的模式也是一种不错（或者无奈）的选择。

“以 Deep Learning 为代表的新型计算模式将引领未来芯片的发展方向”，这一观点基本已经是大家的一个共识。越来越多的玩家会关注能够支持新型计算的芯片，其中很多可能之前完全不在半导体这个圈子，也完全不了解芯片是怎么回事。2017 年我们不时能看到一些对比 CPU，GPU，FPGA 和 ASIC 架构的科普文章，甚至有 10W+ 的阅读量，不难看出大家的热情。

初创公司 • 长长的 list

2017 的 AI 芯片大戏中，主角不仅是巨头，初创公司也都粉墨登场，



戏份一点儿都不逊色。更重要的，在初创公司的“表演”中，中国公司不仅毫不怯场，而且非常出彩。我从 8 月份开始在 github 上维护一个 AI 芯片的列表，既包括大公司的产品，又包括初创公司的情况。到 12 月，这个列表中的信息越来越多，世界范围内的初创公司有 30 多家。而且这个列表还只包含了公开信息，还有很多公司处在 stealth 状态并没有收录。我也听到一个说法，在 AI 芯片领域的初创公司可能超过了 100 家，在 TSMC 排队投片也有 30 家。

不管在什么领域，初创公司都会面临很多风险和不确定性，也可能在成长过程中不断调整和变化。AI 芯片当然也不例外。我们看到，在这一年中，很多公司在不断成长，逐渐明确自己的方向和定位，走的越来越坚实。另一方面，从今年初创公司融资的情况来看，这个领域（也包括更大范围的 AI 概念）也明显出现一些泡沫。有些公司，在没有任何实际东西的情况下，就可以实现“PPT 融资”或者“Paper 融资”。有些公司，重心放在了 PR 上面，功夫都是做给 VC 看的，人称“2VC”公司。面对 AI 这个趋势性机会，有泡沫当然也是正常现象，只是希望这些泡沫不要伤害整个市场的发展。

抛开各种烟雾和泡沫，我们逐渐在这个领域初创公司也看到一些“龙头企业”。比如国内的寒武纪、地平线、深鉴科技和比特大陆，都在 2017

年发布了自己的产品；美国的 Cerebras、Wave Computing、Graphcore 和 Groq（前 GoogleTPU 主要设计者创立），或有雄厚的实力，或有自己特色的文化和比较清晰的产品。在 2017 年，国内也出现一些依托应用开发芯片的 AI 初创公司，这些公司大多以应用牵头研发芯片。我也预期在 2018 年会看到更多这样的情况。当然，很多初创公司并没有公开自己的信息，不排除正在憋大招的可能性。

熟悉半导体产业的朋友可能比较清楚，半导体领域初创公司获得 VC 投资在之前是非常困难的。主要原因是这个产业风险大，门槛高，周期长。但 2017 年，AI 芯片的初创企业却受到了资金追捧。我们可以看看今年的一些公开的融资数据。寒武纪：1 亿美金（估值近 10 亿美金）；深鉴科技：4000 万美金；地平线：近亿美金；；Cerabras：6000 万美金（估值 8.6 亿美金）；Graphcore：5000 万美金。在前面我也提到，当 Nvidia 宣布要开源 DLA 的时候，大家感觉会对初创公司的融资和估值有一定影响。但从结果来看，这种情况并没有出现。在 9 月之后，我们又看到很多初创公司成功融资。而投资者的热情似乎一点都没有减弱，只要有一个新的公司出现，立刻会有很多投资机构蜂拥而至。

为什么传统上不愿意碰半导体产业的投资者现在却对 AI 芯片趋之若鹜呢？这是一个有趣的问题。具体的原因可能有很多方面，整个 AI 领域的投资热潮应该是一个主要原因。如果观察这些投资背后的资本，可以看到很多本身就是 AI 领域很活跃的投资者，甚至本身就是把 AI 作为未来重点的科技巨头，比如 BAT。而传统的投资半导体领域的资金倒是比较谨慎一些。从这个角度来说，这些没有太多半导体背景的资本大量进入芯片领域，是会给大家带来新的机会和视野，还是带来风险和不确定性，还是有待观察的。另外，现在所说的 AI 芯片，一般是指 Deep Learning 加速芯片，相对来说，关键算法简单清晰，优化目标非常明确，很多技术（比如矩阵运算的硬件加速）已有多年的研究基础。而对这种硬件加速器的验证，测试和调试也相对容易。如果不进行精细的优化，硬件部分可以由一

一个较小团队在较短时间完成。这些技术上的特征比较适合初创公司快速尝试。当然，做一个加速芯片（或者 IP）的硬件只是第一步。要真正做出能被市场接受的产品，则需要很多扎实的工作，产品定义，硬件效能，软件工具，系统测试，现场支持等等，一个短板也不能有。虽然大家都很关心投片的时间，但样片出来之后，脏活累活还多着呢。

2018 • 关注什么

对于 2018，我还是非常期待的。作为一名多年从事芯片架构设计多年的工程师，我首先期待看到一些技术上的创新。2017 年我写了不少分析 AI 芯片相关技术的文章，到年末几乎有点审美疲劳了（相信读者也是一样），似乎新鲜东西越来越少。在 2017 年底，有一个叫 Vathys 的初创公司，一下子开了好几个脑洞，全定制的 Asynchronous Logic，等效的时钟可以到 12GHz (28nm 工艺)；High-densitySRAM (1T-SRAM)，片上存储容量可以达到 1.5GB (28nm)；Wireless 3D Stacking，10,000GBit/S @ ~8 fJ/bit。这几项技术要么是目前还停留在学术研究阶段，要么是曾经昙花一现。一个初创公司一下就祭出这几个大招，又是这么高的指标，真有可能实现吗？所以，当 Vathys 的老板发邮件说应该把他们公司加到我做的 AI 芯片 List 里的时候，我开始是婉拒的。不过，换一个角度来看，即使是他们完全在忽悠，也算是击中了 Deep Learning 处理器的痛点。而且这几项技术目前也都有人在研究，在 AI 的热潮和巨大的资金支持下也许真能搞出来也说不定。所以，我还是希望看到他们或者是其它团队能够在这几项技术上取得突破，让我们真正激动一把。说到技术的突破，我们未来（可能要比 2018 年更远）还可以期待看到在存储技术上的突破，以及由新的存储技术带动的架构上的创新，包括 Neuromorphic 这条技术路线。

接下来，当然是巨头们的下一步动作。Google 的 TPU 是否会卖给自己之外的用户，直接和 Nvidia 展开竞争？目前 ONNX 阵营已经形成和 Google 的对峙，Google 作为生态最完整的厂商，推广 TPU 对巩固自

己的领先地位很有意义。Big Five 和 BAT 哪个会学习 Google 榜样直接自研芯片？阿里达摩院的芯片研究会不会从 AI 开始？Intel 能不能如大家所期待的全面爆发？Nvidia 会如何应对来自各方的挑战，是否会展开更专用的加速芯片，而不是仅仅在 GPU 中加个 Tensor Core？高通什么时候在手机芯片中加上硬件加速器？ARM 下一步会怎么走，会不会横扫嵌入端？。。。随便想想就会有很多值得期待的看点。最近我们也看到，为了对抗 Nvidia，AMD 和 Intel 竟然很罕见的宣布合作。而 IBM 在 Power9 上和 Nvidia 深度合作。2018 年也许我们还能看到业界巨头间更多的合纵连横。

初创公司的命运也是 2018 年最大的看点。我在之前的一篇文章中说过“对于 AI 芯片的 startup 来说，2018 年就算不是毕业大考，也至少到了学期末考试了…”。2018 年，大部分初创公司都将会交出第一次测验的结果（芯片），也会开始小批量的试用。相信到时会有比较公平的 Benchmarking 结果出现，“理论上”的指标会被实际的“跑分”结果取代。虽然对于初创公司来说，犯错误是可以容忍的，第一代芯片也不能完全代表公司未来的前景。但是，做芯片需要巨大资源的持续支持，这个阶段掉队可能非常危险。当然，第一次的淘汰对于真正优秀的企业也是最好的机会。我非常期待看到能够在考试中脱颖而出，并跨上新的台阶（或者直接毕业）的同学；或者，会有我们不认识的面孔，突然惊艳出场。另外，2018 年，在 Edge 端会有更多的传统芯片厂商加入竞争，三星，高通，MTK，展讯等等；而在嵌入端 IP 上有绝对优势的 ARM 应该也会有更大的动作，这些都可能会对初创公司的命运产生重大影响。

最后，是变局的可能。从整体上来讲，AI 整体上在 2018 年会怎么发展是一个大家都非常关注的问题。继续高速增长，还是平稳发展，又或者会遇到问题高开低走？不管是哪种情况，AI 芯片必然会受到大势的影响。比较特殊的是，芯片研发的周期大约在 9 到 18 个月左右，这比软件应用的开发和更新周期要长的多。再加上一些滞后效应，芯片的发展很难和算法和应用的发展节奏同步。芯片开发中一个比较可怕的问题就是未来

的不确定性。相对来说，一个可预期的平稳增长的环境是最有利于芯片研发的，可以让芯片设计者能够更好的规划产品和协调资源。另一种变局情况是，算法层面发生巨大的变化，也就是技术上的不确定性。这几年最成功的 AI 算法就是基于神经网络的深度学习。这正是目前 AI 芯片在需求上的基础，也决定了现在大部分 AI 芯片都是以加速这一类算法为目标的。如果基本算法需求发生变化，会对 AI 芯片的设计产生很大的影响。比如，目前已经有一定应用基础的低精度网络，也就是在 inference 中使用非常低的精度，甚至直接使用二值网络。如果这种 Inference 得到广泛应用，现在的芯片架构则可能得要重新考虑。再比如，如果 Hinton 大神的 capsule networks 得到实用，也可能会需要新的芯片架构来支持。毕竟 AI 领域现在发展很快，所以大家也都必须要时刻盯着应用和算法层面最新的进展。我们也要随时问自己下面的问题（来自 Jeff Dean 在 NIPS2017 的演讲）。

If you start an ASIC machine learning accelerator
design today, ...

Starts to get deployed into production in ~2 years

Must remain relevant through ~5 years from now

**Can We See The Future Clearly Enough?
What should we bet on?**

总结

2017 年马上就要过去，在这几年相对“平淡”的半导体领域，AI 芯片让我们小激动了一下。其实可聊的事情很多，以上文字基本上是想到哪写到哪，也都是个人一点点感想，准确的地方，还请各位多多指正，多多包含。

祝各位读者 2018 年万事如意！更要祝各位奋战在 AI 芯片第一线的各位同仁获得成功！

AI 前线声明 | 本文系唐杉博士原创文章，已经授权 InfoQ 公众号转发传播。

作者简介

唐杉博士具有超过 15 年的芯片设计、专用处理器设计和 SoC 架构设计经验，现在 Synopsys AI Lab 负责 AI 芯片架构和相关技术的研究。欢迎关注唐杉博士的公众号 StarryHeavensAbove。



人工智能落地，商业化比技术更重要

作者 | 蔡芳芳

当人们回望2017年的时候，看到的可能全是人工智能。

年末技术盘点里程碑有一大半跟人工智能相关、人工智能企业屡次获得巨额融资、今天人工智能在某个指标上达到甚至超过人类水平、明天人工智能在某个比赛中打败了人类、据分析人工智能可能取代多少人类的工作……

人工智能又一次走到了风口浪尖上，从政府、学术界、企业界、投资界到创业者们，无一不将人工智能视为未来方向；而媒体铺天盖地的报道，更是让人工智能快速占领了每一个普通人的视听。

但是无限风光的背后，又怎么可能没有一点阴影？

“都说自己在做人工智能，其实压根不知道人工智能能做什么。”

“连产品模式都还不清晰，凭什么拿到那么高的估值？”

“说能帮我们解决问题，结果连我们的场景都不清楚。”

“不能大规模商业化、不能帮企业盈利，那为什么要为AI买单？”

高大上的AI技术如何真正落地、带来实际价值，已经成为寻求转型的传统公司、专注于提供AI技术服务的初创企业、业务规模庞大的集团公司共同遭遇的危机。

本文根据平安集团 SMART 科技大会采访内容整理而成。

AI，“看上去很美”

从积极的一面来看，人工智能催生了大量新技术、新企业和新业态，为个人、企业、国家乃至全球提供了新的经济增长点，甚至可能成为第二次“工业革命”。

IDC预计，全球人工智能支出到2020年将达到2758亿人民币，未来五年复合年增长率将超过50%。中国人工智能技术支出将达到325亿元，占全球整体支出的12%。

从消极的一面来看，尽管人工智能开启了一个全新的时代，但也在不断滋生着“泡沫”，吹捧有之，跟风有之，噱头有之。近两年，数十家中美AI创业企业密集倒闭，大量AI创业项目中途夭折，不免让人感慨，人工智能是否只是“看上去很美”？

3年前你都不好意思说你是做人工智能的，而今天你都不好意思说不做人工智能了。

泛人工智能是人工智能火爆之后催生的一种现象，就是把什么东西都叫人工智能。如今没有哪个企业不想拥抱人工智能，但当前人工智能技术的采用程度到底如何？

麻省理工学院Sloan Management Review最近对3000位高管、经理和分析师进行的一项调查显示了一个令人惊讶的结果：目前大多数公司还没有采用人工智能技术，只有不到四分之一的受访者表示已经采用人工智能技术，还有23%的受访者正在进行一些试点项目，而有54%的受访者表示他们还没有开始采用人工智能技术。这与AI前线在年终总结时所做的一份落地情况调查结果相近。

麻省理工学院的调查中也列出了可能阻碍企业采用人工智能的几大障碍：

对于已经理解并采用人工智能的组织（“领导者”），人才缺口、竞争激烈的投资和对安全的担忧是他们的主要障碍。

与此同时，尚未采用人工智能（“被动”）的企业认为需要识别业务应用场景、缺乏管理支持、技术能力有限是其主要挑战。

AI 落地困境

当前的人工智能实际上是数据驱动的智能，也就是基于大数据的智能。中国科学院深圳先进研究院首席科学家须成忠教授认为这一波人工智能涉及ABCD四个概念，A算法，B大数据，C计算平台（如云计算平台），D领域知识，技术必须跟领域结合。技术不结合场景，就只是一个技术而已。

目前AI在医疗、金融、交通等多个领域皆有不同程度的渗透，虽然行业不同，但AI落地时遇到的困境却多有类似，其中场景和数据是最常被提到的两项。

为什么企业要为AI买单？

据平安医疗健康管理股份有限公司首席技术执行官穆强介绍，目前AI在平安医疗健康的应用主要分为面向病患和面向医护人员两种类型。

- 面向病患：利用AI技术对患者分层，通过患者的就医习惯和行为画像定义风险等级，再匹配需要的医疗水平，也叫做医疗能力分层或者患者需求分层，从而把医疗的供给侧和需求侧匹配起来、协调医疗资源。
- 面向医护人员：用AI给医疗行业赋能，借助AI为处方点评、临床决策等提供决策建议，帮助医生更好地决策判断。另一点是运用医学知识图谱和大数据的方法，调整综合医疗费用结构。目前的总费用中，药物占比过高而医生的价值占比太低，因此需要在总费用不变的前提下调整医生价值与药费的占比，使医生发挥的价值与价格不背离。

平安尝试在医疗场景落地AI的过程中，既会使用公司内部的技术方案，也会与第三方公司合作。商业合作不是公益项目，企业最关心的自然是这项技术到底能为自己解决什么问题？能带来什么价值？而这可能也是AI技术服务商在尝试向行业输出技术前，最需要想明白的问题。

穆强指出，AI初创公司想要进入医疗垂直领域，最大的痛点是买单方，也就是谁为他们的AI技术买单的问题。“企业经营不外乎开源节流。如果想把AI技术应用到产品中，首先要找到AI能给企业经营带来什么价值，是能帮企业开源还是节流，还是抑制风险，还是提高医疗水平，总得帮人干点什么，所有的这些最终都要转化成商业价值。只要这项AI技术能够帮我们提高经营能力，我们自然愿意为之付费。”

错把商业项目当成研究项目

须成忠教授是中国科学院深圳先进技术研究院首席科学家，他所带领的团队开发了“先进云”平台，平台之上融合了大数据、机器学习等技术，在智能交通、医疗健康、城市信息等多个领域都有落地应用。

须成忠教授告诉AI前线，智慧交通即利用人工智能和大数据使城市交通变得更智能，其终极目标是通过城市大脑做预测，并能持续学习进而反馈控制，但目前暂处于起步阶段。“阿里做城市大脑，说将城市拥堵改善了10%，其实这10%从科学研究角度来说误差都要拿掉了，可以说是微不足道，而且它是基于一个小区的小范围实验。”

目前中科院深圳先进技术研究院在交通领域已经落地的成果主要基于中科院收集的城市所有浮动车（出租、公交等非固定权的车辆）数据，包括三类服务：面向政府的服务，如城市精细化管理；面向公众的服务，如深圳市的公交在手软件，能够综合实时路况、天气情况预测当前公交距离本站还有多久；将数据通过API方式开放给第三方企业（如腾讯、广电集团、酷米客等）做更多的应用和服务，每天数据API访问量达到500多万次。

在加入中科院之前，须成忠教授在国外也承担过不少企业合作项目。

在他看来，做技术研究的人最容易犯的错误就是太过在意“研究”，错把商业合作当作科研项目。

须成忠教授指出：“研究院研发的产品原型充其量是证明你有这个技术能力，但还是需要针对用户的具体需求来修改和定制开发。我们以前把项目当作研究项目来做，并没有解决企业的刚性需求，研究院经常会遇到这样的问题。结果导致我们做的技术，用户并不是特别感兴趣。”

须成忠教授举了一个例子，假设客户要做出租车实时报表，他们原来平时做一次报表要花两个多小时，客户的需求是实时化。利用大数据处理将报表处理时间降到一分半钟后，客户就很满意了。至于一分半钟和一分钟的差距，就属于研究问题了，这个问题客户可能并不感兴趣。虽然从研究的角度来说，时间缩短了50%，但对企业来说价值不大。

“技术输出首先要搞清楚解决企业问题和研究项目的差别。只要技术好，然后能够钻进去跟客户一起做一些面向客户需求的定制化开发，肯定能够生存下来。”

无数据不AI，无场景不AI

传统网络设备商思科早已开始数字化转型，同时也期望能够通过技术帮助行业升级。全球很多流量都跑在思科的网络设备上，“如今大家都在关注AI，但其实最核心的是底层数据。”思科大中华区副总裁、大客户事业部经理海广跃告诉AI前线，而底层数据正是网络设备商思科的强项。

思科尝试将网络设备变成传感器，从底层基础设施开始引入AI、大数据，帮助各行各业进行技术升级和转型。据介绍，目前思科的数字化转型支持已经覆盖了11个行业，包括互联网公司、运营商、金融和娱乐行业，医疗和教育行业相对受AI影响慢一些，石油石化、电力系统最慢。

不同行业AI落地速度快慢，有很大的因素与数据有关。由于这一波AI对数据的重度依赖，应用首先得有足够的数据，而这可能正是很多传统行业所缺失的。即使有了数据，数据是否正确、是否足够有代表性、如何准确标注又成了另一个老大难。“种瓜得瓜，种豆得豆”，机器学习的成果取

决于数据的质量，错误的数据只会让训练结果出现严重偏差。

海广跃指出，AI最核心的价值在于数据和场景，而目前很多传统行业难以应用AI或者应用效果不佳一般也是数据或场景的问题。

- 数据方面的问题包括：
- 数据太少甚至没有，比如电力行业、工厂车间；
- 数据实时性，以自动驾驶为例，车本身是要和周围环境交互，包括公路情况、天气情况、行人聚集情况，需要实时判断并决策，网络存在时延，如果不能实时响应就可能会出现事故；
- 数据处理需要借助领域知识和经验。先找到特征值并进行标注，经过训练之后才能用来做决策或预测，不懂行业的数据科学家无法做到这一点，比如医院X光片、工厂设备故障检测等。而没有特征值就没办法应用AI，只有懂行业、懂场景的人与数据科学家配合才能真正将AI用起来。

软通动力集团执行副总裁康燕文也向AI前线表达了相同的观点。

在语音识别、多国语言翻译、音视频结合方面，软通动力集团与不少大公司有合作。软通负责提供语音数据和产品测试服务，辅助验证语音产品的可用性，而在这样的合作中数据正确性显得尤其重要。

康燕文指出：“在对AI产品进行测试时，如果数据给的不对就沒用了。比如展厅机器人，有很多人进来参观跟它对话，机器人有知识库，它需要从知识库中找到你问的问题，它才能回答，你问的问题不对或者不在知识库中就没办法回答。”

为了升级而升级

AI如此火爆，不乏“拿着锤子找钉子”的初创企业，也有不少还没想清楚为什么要升级就想先把AI“用”起来的传统企业。“这不太现实。所有的技术升级都是业务驱动，不能凭空自己造，也不能为了升级而升级。”氪信CreditX首席科学家周春英告诉AI前线。

氪信是一家专注于运用机器学习与大数据技术解决金融信用风险问题

的服务商，目前已经与数百家金融机构达成合作。

营销和风控是目前机器学习在金融领域的主要应用点。营销方面，主要就是解决不同的客户和不同产品的匹配问题，以实现精准推荐。在这方面应用机器学习更多是一种“升级”，由于整合的用户数据和产品数据更多，相比简单的LR，推荐成功率有所提升。风控则相对复杂，涉及反欺诈、信用评分、行为监控等。其中反欺诈分为两个维度，一是基于个人数据，和营销类似，从简单的评分卡转为机器学习，升级数据处理和模型技术、提升效率；另一个是为机构建立全套体系，包括机器学习的模型、算法、训练等，这不是一次性的工作，需要不断迭代。

周春英认为一开始寻找切入点是最难的，“要考虑怎样才能快速推进，又能落地上线使用，还能够不断迭代升级”。大家都知道要技术升级，问题是怎麽找到需要升级的那个实际业务问题，如果平白无故地去做技术升级，没有落地的实际效果很难推动，也难以形成不断迭代的过程。

氪信对此的解决方案是合作，在一开始会花比较多的时间去和客户、业务负责人交流，共同寻找切入点。周春英补充道：“找到切入点之后其实也不是一下子从0到100，而是需要花一段时间（比如几年）才能出来成果。”

另外，不同客户的需求不一样，这就需要对解决方案和产品的抽象能力，既不能过度设计，也不能不抽象。在与客户的合作中，既要保证高效率，又要满足其个性化需求，因此方案和产品的抽象设计是对于AI技术服务商的另一项挑战。

传统行业对新技术不够开放

AI技术不是单点技术，它需要形成一个体系，因此免不了需要巨大的投入，从底层硬件到数据再到专业人才，缺一不可。对于金融机构来说，硬件和系统花钱买不难，但AI人才培养比较难。

周春英认为这是专注技术的AI初创企业的机会。术业有专攻，技术的问题可以交给懂技术的人，但寻找业务场景离不开良好的合作。合作中可

能遭遇的问题就是传统行业不够开放。虽然金融领域有其特殊性，包括对信息安全、数据安全都有特殊的需求，但是如果金融机构越来越封闭，对于新技术的应用只会有不好的影响。

“对于新技术、新方式，或者新的技术方向，传统行业要有一个更开放的态度，包括数据的使用，这样实践或者是落地可能会更快一些。”周春英说，这或许也是很多AI公司内心的呼喊。

如何将AI技术通用化？

平安科技是平安集团32个子公司之一，为整个平安集团提供技术服务，此外也给国内的众多银行和保险公司提供服务。过去几年，平安科技一直在进行AI的研发，目前在大数据平台的基础上，引入了面部识别、语音识别、视觉识别、行动识别、情感情绪识别等技术，进行自动决策。

以贷款申请为例，平安科技利用人脸识别匹配客户身份、基于各项数据智能分析客户的风险状况，3分钟之内就可以在网上给客户派发贷款。

平安科技大数据高级产品总监王建宗告诉AI前线，平安集团在AI时代转型的难点不在于做不做AI。对于先在公司内部落地AI之后，再开始尝试将AI技术对外输出的平安科技，面临的难题是：如何将AI技术与自身业务剥离，使其通用化、标准化、产品化？

王建宗指出：“首先，每个公司的数据不一样，数据标准参差不齐，平安的数据是平安的一套，但另外一个公司可能是另外一套数据，中间涉及数据融合、数据归集化、数据交叉，这是一个挑战。第二个挑战来自于业务流程的差异性，保险行业每个公司的保险流程都不一样，又涉及流程融合。”因此平安科技在帮助合作伙伴落地AI的时候，既会提供模型和算法，也会提供能对模型和算法进行闭环训练的系统，使数据和规则能够快速通过系统优化，并应用到实际业务中。

破局之道

看清：AI不是银弹

当大数据将AI推向“技术炒作”舞台的正中央，带来的一个令人混淆的结果就是，突然间人人都号称自己做的是“AI”，实际上却是徒有其名；人人都以为自己需要“AI”，却不知道AI到底能为自己解决什么问题。

在硅谷人工智能研究院创始人Piero Scaruffi教授看来，以上问题不只出现在中国，在全球任何地方都一样。

AI真正应用于商业场景通常会面临两个问题：第一，有些商业人士可能擅长传统的科技思维，但是对AI了解不多；第二，很多优秀的软件工程师从未学习过AI，他们学的是其他领域的技术，AI不是他们擅长的东西。对于工程师来说，有计划地学习一些课程就能解决问题，但对于商业人士则非常困难。

面对这些问题，企业决策者和技术管理人员首先应该明确一点：AI不是万能药。当问题已经有解决方案时，可能没必要采用AI，否则反而是在把工作复杂化，或者变成自己给自己制造问题。

Scaruffi教授指出，没有AI商业照样可以运转，现在大多数AI仅是优化程序，比如机器人。为什么买机器人？为什么用机器人代替原来的工作？因为机器人更便宜，可以降低成本。但是这些所谓的机器人一般都非常简单。

“当有人说中国每年制造1000万台机器人时，你应该问问这是什么机器人。一天到晚都在做简单机械工作的机器人并不是AI，但是它们的确省钱。这可以称作自动化，但已经是存在很久的技术了。而真正的AI，比如说可以回答问题、实时监测的机器人，一般非常贵。”

正如前面所说，这一波AI是基于数据的AI，因此AI应用到实际问题中离不开数据。没数据就不能训练神经网络，不管计算能力多强、速度多快。无数据无AI，如果没有就需要先搞到数据。如果有数据，AI可能是有用的。但在一些情况下其实大数据技术就可以解决企业的问题，未必需要用到AI。企业必须能够分清楚哪些问题需要用AI来解决，哪些问题用已有的其他技术就可以了。

康燕文认为，应用AI之前首先对AI的认识要清楚，要知道AI到底可以

做什么、局限在哪里。比如制造业要用AI识别成品存在的缺陷，需要先积累大量数据，让机器去学习，才能开始识别。工人师傅需要积累经验，机器同样需要积累数据，获取数据、积累数据都需要一个过程。康燕文还补充了一点，即新技术需要测试环境，不能一下就拿来用，出错的风险太大，需要先做一些实验。

须成忠表示，人们应该要修正自己对于人工智能的期望。虽然人工智能目前在封闭式环境中确实有了一些很好的突破，但是在开放环境中还有很长的路要走。另外，人工智能是无法在一朝一夕突然就能够解决所有问题的。

“我们不能也不要头脑过热，过一段时间发现人工智能除了下棋打牌其他都做得差强人意。人脸识别需要在标准场景下，语音识别在开放环境还有问题。更何况数据清理和转换也要花掉不少精力。如果大家不能修正对它的期望，到时会带来更多失望，但这也是一个过程。”须成忠说。

合作：寻找AI落地场景

有人将AI落地问题比作“道”和“术”的问题。做AI工程的人，最在意的是这个算法怎么样，够不够快，这是“术”的角度；做业务咨询的人更考虑你有什么问题，我要帮你解决问题，我用什么你别管，我有创意，最后成了就行，这是“道”的角度。

那么当掌握了AI技术或有了科研成果，怎么跟企业需求对接？如何在企业的业务场景中找到应用新技术的切入点？

海广跃认为，首先得将业务或场景流程化，才能找到存在问题、可以应用AI的环节，接下来再考虑怎么以AI为工具去解决流程中存在的问题。另一点则是要站在未来看现在，看未来会是什么样的使用习惯，找到未来的场景，就能够知道该做什么样的产品。这是另一个如何创造新业务的问题。

周春英则表示，寻找应用场景的切入点其实没有什么窍门，最实际的就是做技术的人必须懂业务，而方法总结起来也简单，就是多沟通、多学

习。通过经常和不同的客户交流和沟通来学习业务，先从客户的出发点考虑问题，然后再转化为用什么技术去解决它。

“要帮助客户解决问题，既要懂业务也要懂技术。光懂技术跟客户谈不到一起去。弄懂业务之后，再想怎么把业务的痛点转化为技术的解决方案。接下来把这个方案讲通讲明白，这个事情怎么做，实际的痛点是什么，什么技术能解决，出什么样的解决方案，这个就是我们实际项目过程中的经验。”周春英说，“跨学科能力是技术人必备硬技能。”

须成忠教授也认为切入点需要双方共同碰撞出来，而“拿着锤子找钉子”的做法目前并不是很可行，因为通用的一般性技术、一般性系统很难找到，也很难有市场。

“软件行业开源那么多，但实际上开源只是证明你的能力，最后还是需要做定制开发。至于说怎么找到用户的需求，首先要让用户知道你能做什么东西，有什么样的能力，然后再根据用户的需求去做深做透。这是双方互相碰撞的一个过程。需要双方沟通后，弄清楚我有什么问题，你有什么办法能解决我的问题才能确定目标。”须成忠教授说。

基于平安医疗健康的业务场景，穆强给AI前线列举了AI在医疗领域可以切入的角度。

保险就是管理可控风险，怎么让风险变得可控？其实就是两个问题，怎么收钱以及怎么花钱。这两个方面规则的定义恰好是如今基于大数据的AI所擅长的——基于客观事实数据和规律，将主观判断变客观，从而更合理地制定规则和策略。

医保风险控制包括三点，一是防止欺诈滥用浪费，二是降低疾病发生风险，三是提高被保人的健康水平，而这些都是综合降费的办法。“对于保险公司来说，1%的重疾和19%的慢病会花掉70%的医保费用，那么这20%的人就是保险公司重点管理的对象。怎么把这些人挑出来？怎么干预这些人的行为，对他们进行个性化管理？这对我们来说投入最小、见效最快，如果AI可以解决这些问题，能够帮我们降低成本，就有落地的价值。”穆强说。

渐进：平衡AI的投入产出

目前不乏有些企业辛辛苦苦捣鼓了AI，但效率却得不到相应价值的提升。那这些投入只能算是打水漂了吗？企业应该怎么评估AI技术能带来的实用价值？又该如何衡量应用AI需要投入的人力成本、时间成本？

AI是大势所趋，随着未来5G、物联网的发展升级，数据只会越来越多。以后谁能利用好数据，谁就会在业界领先，而不会用数据可能就会被淘汰，所以AI升级肯定是要做的。但现阶段，逐步迭代可能是比一步到位更靠谱的选择。

周春英表示，AI本身就是比较重投入的，但是它也会有更长远的影响，所以不能特别短视于此时此刻的回报上。但不同企业的需求不同，不能一概而论。周春英告诉AI前线：“我比较倾向于，AI技术升级不要想着一蹴而就，一步就到深度学习，而是逐步迭代，先有一个整体的规划，然后一步一步慢慢取得成果。这样能够比较好地平衡投入和产出，对于一些小型金融机构来说确实是这种做法比较好。”

康燕文认为，不能把新技术拿来就用，需要一个实验的过程，循序渐进，而不是说一下子投个大钱就想得到巨大突破，这样才能避免出现投入很多却得不到预期结果的问题。“人工智能是大趋势，现阶段企业更应该赶快去试，越不去试，越搞不清楚，等试了之后慢慢找到对的方法和方向，再一步一步往前走，自然就会越来越清楚。”康燕文说。

方向：民主化还是定制化？

最近不少企业都在谈AI民主化、AI普及化、AI服务化，或者AI for all，其实不管名字怎么叫，其核心就是一点，希望能够降低人工智能的门槛，让非专家也能使用AI，这也是很多人对于未来AI的期望。

海广跃是这个观点的支持者，他认为推进AI民主化，才能推动AI在各行各业的应用普及。

海广跃说：“当AI高高在上的时候，传统行业懂场景不懂技术的人，与懂技术不懂场景的专家之间有鸿沟，AI和数据、AI和应用场景、AI和

知识之间也同样存在鸿沟。”

至于如何跨越这个鸿沟，海广跃也提出了他的想法。一是构建一个各行业通用且低价的AI平台，使不同行业的人都能很方便地使用AI，在平台之上运行业务而非研究技术。但他也坦承，难度比较大，因为很难用一个AI平台满足各行各业的不同需求。偏底层的AI基础平台或许可行。

另一点就是纵向打通，将一个垂直行业场景从上到下全部打通，把这个场景做好了，再做另外一个场景，先对某一个企业产生价值之后再继续做加法。

须成忠教授则认为垂直领域的深挖更有潜力。他告诉AI前线：“通用的一般性技术不一定有市场，充其量就是开源，想要发展就一定要把这个通用技术应用到垂直领域，然后针对这个垂直领域来做定制化的开发。将来的人工智能不管发展到了哪个阶段，一定不会像人那样什么都懂、什么都会，它一定是这个软件和系统专注一个业务，另一个专注另一个业务，是跟领域相关的。未来AI会越来越偏向跟应用接轨，越来越多地与应用相关。”

AI 的未来

有人说人工智能是未来已来，也有不少人认为人工智能的蜜月期会终结于2018年。

AI离我们是近还是远？淘宝搜索、银行人脸支付、小米智能音箱，无处不是人工智能，对于普通消费者来说，AI似乎很近。但对于企业和技术人员来说，不是所有问题都像消费级产品一样容易解决。人工智能所有涵盖的技术纷繁复杂，怎么从中找到合适的方法来解决自己的问题？怎么得到数据？怎么清洗数据？甚至一开始怎么找到应该解决的问题？这些都是企业想接近AI却又觉得AI“遥不可及”的路障。但这些也是未来要让人工智能真正扎根于工业界、真正改变每个人的生活所必须解决的问题。

人工智能落地难，却是这项技术真正变成“新电力”般存在的必经之路。对于传统行业公司与AI技术初创公司来说，都是挑战与机遇并存。

原百度研究院院长林元庆从百度离职后，创办了自己的公司AIbee，他的目标是利用AI深度赋能传统行业，实现产业升级。而人工智能著名学者吴恩达也于今年12月宣布成立Landing.ai，立足于解决AI转型问题，第一站是制造业。

从国家政策来看，12月工信部发布了《促进新一代人工智能产业发展三年行动计划》，从培育智能产品、突破核心技术、深化发展智能制造、构建支撑体系和保障措施等方面详细规划了AI在未来三年的重点发展方向和目标，足以看出国家对AI产业化的重视。未来AI企业也将从拼技术逐步转移到拼行业、拼落地。

2018年，人工智能的泡沫是否真的会破灭，我们不得而知，但过度期望带来的泡沫被戳破未必是件坏事。

接下来一段时间，人工智能落地势必会是业界最为关心的话题，而人工智能落地无非是：弄脏手、扎进去、踏实干。



DeepMind 的 2017 年：历程回顾

作者 | Demis Hassabis 等

译者 | 核子可乐

今年 7 月，世界头号棋手柯杰在获得 20 连胜后作出发言。而在两个月之前，他在中国乌镇举办的围棋峰会上与 AlphaGo 展开了对弈。

他表示，“在与 AlphaGo 进行对弈之后，我从根本上转变了自己对比赛的审视角度，现在我意识到这样的思维转变确实很有帮助。我希望所有围棋选手都能够考虑到 AlphaGo 对于围棋竞赛思维的理解方式，这确实极具现实意义。虽然我办理了，但我发现了围棋当中蕴藏的巨大可能性，而围棋本身也在不断进步。”

乌镇围棋峰会是一场为期五天的盛会，其中包括双人赛、团体赛以及一对一柯杰挑战赛等

柯杰是这场比赛的绝对中心，我们对他的结论也感到荣幸万分。他的发言给我们带来启发，因为这暗示着未来的社会将能够利用 AI 作为发现工具，用以发现新知识并增加人类对世界的理解能力。特别是在机器辅助科学方面，我们希望人工智能系统能够帮助我们有效应对气候变化与药物



发现层面的挑战，并立足发现复杂的新型材料或帮助缓解医疗系统压力等领域取得进展。

这种社会效益层面的发展潜力正是我们设立 DeepMind 的原因——我们很高兴能够在解决基础科学挑战以及人工智能的安全与道德等方面不断获得喜人的进展。

我们在 DeepMind 所采取的方法受到了神经科学的启发，其帮助我们在构图、推理、记忆以及学习等关键性领域取得了进展。以想象力为例：这种独特的才能在我们的日常生活中发挥着关键性作用，使我们能够面向未来进行规划与推理。但在另一方面，从传统角度讲我们几乎无法设想计算机会拥有这样的能力。我们为了解决这个问题而持续努力，并于今年引入了起想象力增强代理，其可从环境中提取相关信息以便作出未来规划。

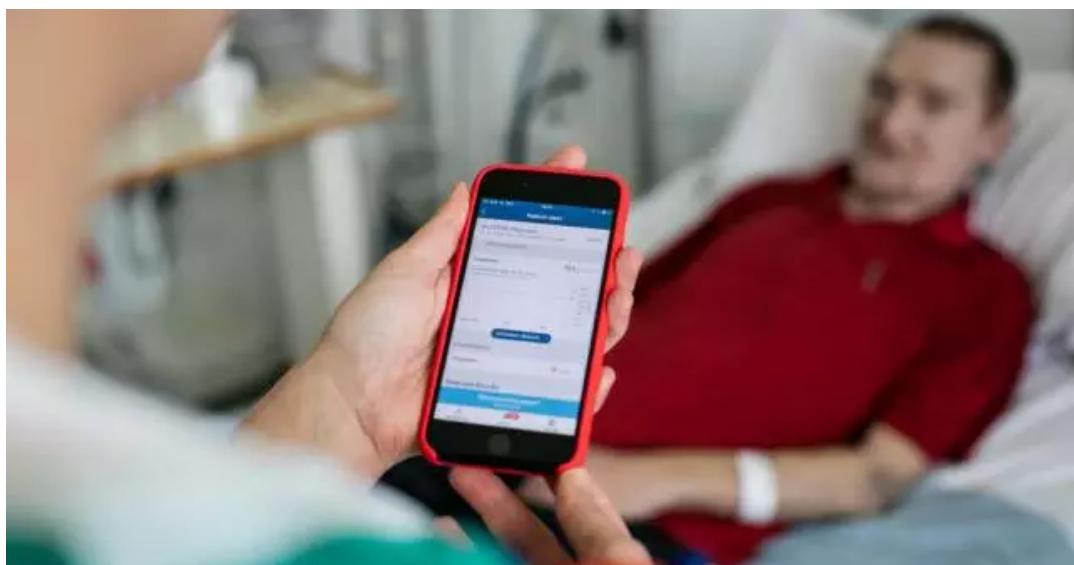
当我们训练神经网络以控制模拟环境中各类经过简化的肢体活动时，这一同样源自神经科学启发的方案也创造出目前我们最为流行的工作演示之一。这种精密的电机控制能力正是物理智能的一大实现标志，亦是我们研究计划中的重要组成部分。尽管由此产生的活动有时显得比较疯狂，但其同时也代表着一种惊人的成功——外加一点娱乐性。

我们都知道，技术不可能完全脱离价值存在。我们也很清楚，除了在基础科学研究方面取得进展，我们还需要在道德与社会影响层面负起责任。

另外，我们在创成式模型方面也取得了进展。就在一年多之前，我们推出了 WaveNet，这是一套用于生成原始音频波形的深层神经网络，能够产生比现有技术更好、更逼真的语音成果。那个时候，这套模型还只是原型设计，而且对普通消费产品而言其计算资源密度过高。但在过去十二个月中，我们的团队创造出一套速度可达其 1000 倍的新模型。今年 10 月，我们宣布这套新的并发 WaveNet 已经在现实世界中得到正式采用，且负责生成谷歌助手的美语与日语发音。

作为我们研发工作的一大实例，这些努力使得 AI 系统的构建、训练与优化变得更加轻松。今年我们还将利用分布式强化学习、基于填充的神经网络训练以及新型神经架构搜索方法等其它技术，进一步改进系统的构建难度、精确性以及优化水平。我们还投入了大量时间以创造新的以及更具挑战的系统测试环境，包括我们正与暴雪公司合作开发《星际争霸 2》的人工智能选手。

正如前文所言，考虑到我们需要为技术开发工作所负起的道德与社会影响责任，我们也一直在探索新的方法以理解并解释人工智能系统的工作方式。正因为如此，我们才建立起一支成熟的技术安全团队，其负责开发各类实现方法以确保未来的系统将继续处于人类的控制之下。



我们现在与四家国家医疗服务（简称 NHS）信托机构开展合作

今年 10 月，我们又建立起 DeepMind Ethics & Society 这一研究单位，旨在帮助我们探索并理解 AI 技术所带来的实际影响，最终实现与预期相符的社会收益。我们的研究工作得到了相关领域著名专家的帮助——包括哲学家 Nick Bostrom、气候变化专家 Christiana Figueres、首席研究员 James Manyika 以及经济学家 Diane Coyle 与 Jeffrey Sachs 的指导。

AI 的设计必须立足于社会的优先层级与关注点，也正因为如此，我们才会积极与各合作伙伴组织开展关于人工智能方案设计与部署的对话。举例来说，算法正义联盟负责人 Joy Buolamwini 与来自《Article 36》、《人权观察》以及英国武装部队的专家们都参加了我们的 Wired Live 会议，在这里讨论算法中的偏见以及如何限制自主武器的致命性能力。正如我们在今年就经常提到的，这些问题极为重要且影响范围非常广泛，绝对不容忽视。

也正因为如此，我们才需要立足各 AI 企业之内与之外的更多新空间以组织更多能够预测并指导人工智能技术影响的讨论。例如，我们今年组织的 Partnership on AI（简称 PAI）会议即将各行业竞争方、学术界与民间社会代表汇聚在一起，共同讨论与人工智能相关的关键性道德问题。在过去一年中，PAI 迎来了 43 名新的非营利性与营利性组织成员，外加新任执行董事 Terah Lyons。在未来几个月中，我们期待着能够与其进一步开展合作，从而研究算法中的偏见与歧视、机器学习对自动化以及劳动力的影响等广泛的议题。

我们对自己在 2017 年年内获得的各项进展感到自豪，但也很清楚未来还有更漫长的发展道路要走。

我们也相信，我们打造出的技术成果对于实现社会效益拥有极为重要的意义——这一切都有可能在健康与能源等方面为现实世界带来巨大回报。今年，我们与 NHS 医疗信托机构建立起两项新的合作关系。其中部署的 Streams 应用将负责利用数字化技术支持 NHS 的临床诊疗。我们还成为领先研究机构联盟中的一分子，并发起一项开创性研究以确定尖端机器学习技术是否有助于改善乳腺癌的检测效果。

与此同时，我们也一直在努力监督自身在医疗卫生方面的工作成效。我们将与英国皇家自由医院合作过程当中，由英国信息委员会专员记录下的经验教训进行了整理。此外，DeepMind Health 的独立审查人也发布了其第一份与我们工作内容相关的年度公开报告。他们的审查将使我们的工作成效更上一层楼。我们在患者及公众参与度方面也作出了重大改进，包括为患者及护理人员召开研讨会，同时亦探索在自身系统之内通过技术手段建立信任——例如计划发布一款可进行审计数据验证的开源工具。

我们对自己在 2017 年年内获得的各项进展感到自豪，但也很清楚未来还有更漫长的发展道路要走。

在乌镇围棋赛的五个月之后，我们宣布 AlphaGO 正式退役，并通过《自然》杂志上发表的第四篇论文介绍了其替代版本——AlphaGo Zero。AlphaGo Zero 的核心特性，在于其不需要任何人类知识的引导。在数百万场对弈当中，这套系统得以逐渐从零开始学习围棋，并在短短数天内就积累起人类沿袭数千年知识。更重要的是，这种重新探索亦提示了围棋中的一些非常规策略，为这种古老的竞技项目带来了新生。

我们的信念是，人工智能必将有能力作为科学工具以及人类智能的放大器，在更多其它领域为复杂问题的处理作出贡献。AlphaGo 团队已经马不停蹄地开启下一轮挑战，希望他们在算法创建的道路上高歌猛进！



阿里巴巴年度技术总结： 人工智能在搜索的应用和实践

作者 | 欧文武

以深度学习为代表的人工智能在图像、语音和 NLP 领域带来了突破性的进展，在信息检索和个性化领域近几年也有不少公开文献，比如 wide & deep 实现了深度模型和浅层模型的结合，dssm 用于计算语义相关性，deepfm 增加了特征组合的能力，deep CF 用深度学习实现协同过滤，rnn recommender 采用行为序列预估实现个性化推荐等。工业级的信息检索或个性化系统是一个复杂的系统工程，深度学习的工业级应用需要具备三个条件：强大的系统计算能力，优秀的模型设计能力和合适的应用场景，我们梳理了过去一年多搜索在深度学习方向上的探索，概要的介绍了我们在深度学习系统、深度学习算法和搜索应用落地的进展和思考，希望对大家有所启发。

深度学习在搜索的应用概括起来包括 4 个方面：

首先是系统，强大的深度学习训练平台和在线预测系统是深度学习应用的必要条件，目前我们的离线深度学习框架、在线深度学习框架和在线预测框架统一到 tf，并实现了日志处理，特征抽取，模型训练和在线服务部署端到端的流程，极大提升了算法迭代效率；

其次是搜索应用，包括智能交互，语义搜索，智能匹配和智能决策四个技术方向，这四个方向的协同创新实现了搜索全链路的深度学习技术升级，并具备从传统的单场景单目标优化到多场景多目标联合优化的能力；

再次是在性能优化上做的工作，包括模型压缩、量化、低秩分解再到二值网络，大量的技术调研和论证，为未来提高深度模型预测性能和软硬件协同优化做了很好的技术铺垫；

最后是排序平台化，实现了 PC 商品搜索、无线商品搜索、店铺内搜索搜索和店铺搜索的搜索服务统一，通过特征和模型复用，实现了多条业务线技术的快速升级。下面我会简要的概括下在四个方向上取得的主要进展和背后的思考。

下面是搜索系统和算法的简图。系统包括：

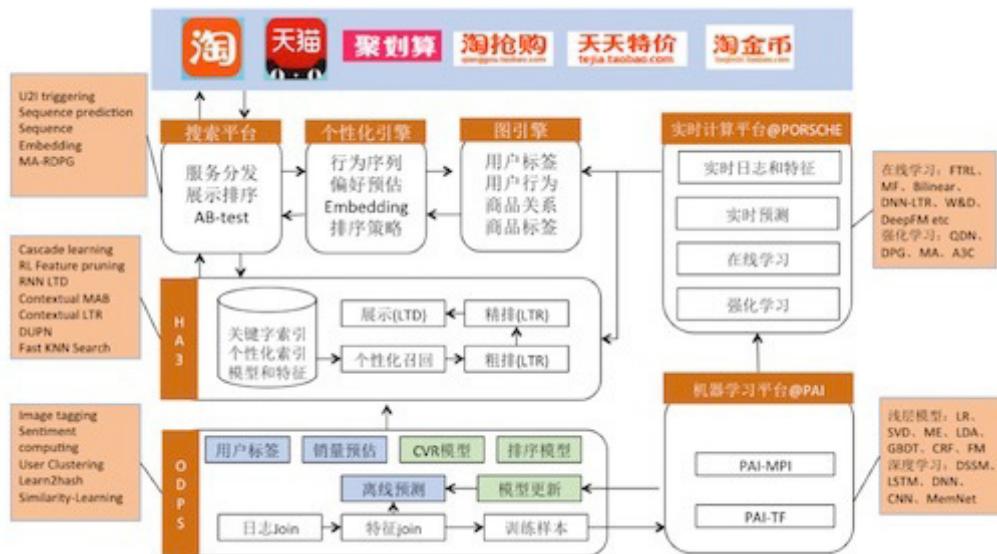
a. 离线数据平台 ODPS，负责离线日志 join、特征抽取和离线模型预估产出排序特征，时效性不强的特征都是通过离线数据平台产出的，比如用户性别标签，商品关键字等；

b. 离线机器学习平台 PAI，底层是主流的 parameter server 和 TF 深度学习框架，平台实现了大部分机器学习算法模型的并行训练和预测，在搜索应用中主要作用是离线模型训练产出离线排序特征模型；

c. 流式计算和在线学习平台 Porsche，流式计算是基于 blink 负责实时日志解析和特征 join 生成实时排序特征，在线学习和离线学习底层框架可以相同，差别主要是依赖数据源和部分优化方法不同，由于用户行为和市场环境变化快，流式计算和在线学习在搜索应用非常广泛，并积累了不少在线学习和强化学习算法；

d. 在线服务平台，包括引擎、排序服务和搜索平台组成，负责在线的服务分发、索引查询、排序服务和结果合并等功能，搜索的排序策略、

相关性、个性化等模型主要通过在线预测服务生效。经过多年发展我们已经具备了非常完善商品搜索排序算法体系，包括知识图谱、分词、tagging、类目预测、意图预测、拼写纠错、query 推荐、query 语义改写、相关性、商品标签、商品质量、店铺分层、用户 profile、用户偏好、用户感知、召回策略、个性化模型、多样性策略、异构服务混排策略、多目标联合优化策略、多场景联合排序策略等，并平台化的方式赋能相关业务团队。



搜索系统和算法简图

系统进展包括机器学习平台和在线预测平台

机器学习平台。搜索训练样本主要来自用户行为，由于用户行为是流式数据，适合做在线深度学习，但当模型参数非常庞大需要海量的样本时在线学习需要很长的时间才能收敛，这时一般是先做离线预训练再结合增量或在线学习，另外有些模型离线预训练后在线只需要对接近输出层的网络做 fine-tuning。搜索在实际应用的有离线机器学习平台 PAI 和在线机器学习平台 Porsche，两个平台深度学习框架目前都统一到了 tf-pai，tf-pai 对原生 tf 做了一些优化，比如底层数通，稀疏参数存储、优化方法、GPU 显存优化等，比原生 tf 训练深度有较大的提升，训练上千万样本和

上百亿参数的深度模型毫无压力。虽然 Porsche 和 PAI 都支持 GPU，但在搜索应用中 CPU 依然是主流，GPU 应用比较少，原因主要是个性化相对图像或语音简单，特征抽取网络比较浅，维度相对较低，GPU 的稠密矩阵计算能力得不到充分发挥，同时离在线混布后流量低谷期间腾出了大量的在线服务闲置 CPU，把临时闲置的 CPU 利用起来做深度学习训练是一个非常好的思路。

在线预估 RTP，搜索排序算分服务。由于每次搜索请求有上千个商品需要计算排序分数，深度模型应用对 RTP 服务的压力是非常大的，RTP 通过采用异构计算，计算算子化和模型分片等方式解决了深度模型 inference 计算和存储问题，深度模型用 GPU，浅层模型用 CPU，今年双 11 期间搜索 RTP 服务用到了 550 张 GPU 卡。另外，RTP 还实现了离线 / 在线训练模型 / 数据和在线预测服务部署的无缝衔接，算法训练好的模型或数据可以很轻松的部署都在线服务，提升了算法迭代效率。

算法包括智能交互、语义搜索、智能匹配和搜索策略四个方向

智能交互。商品搜索就是带交互的商品推荐，用户通过关键字输入搜索意图，引擎返回和搜索意图匹配的个性化推荐结果，好的交互技术能够帮助到用户更好的使用搜索引擎，目前搜索的交互主要是主动关键字输入和关键字推荐，比如搜索框中的默认查询词和搜索结果中的文字链等，推荐引擎根据用户搜索历史、上下文、行为和状态推荐关键字。

和商品推荐的区别是，关键字推荐是搜索链路的中间环节，关键字推荐的收益除了关键字的点击行为外，还需要考虑对整个购物链路的影响，包括在推荐关键字的后续行为中是否有商品点击、加购和成交或跳转到另外一个关键字的后继行为，这是一个典型的强化学习问题，action 是推荐的关键字候选集合，状态是用户当前搜索关键词、上下文等，收益是搜索引导的成交。除了被动的关键字推荐，我们也在思考搜索中更加主动的交互方式，能够做到像导购员一样的双向互动，主动询问用户需求，挑选个性化的商品和给出个性化的推荐理由，目前阿里搜索团队已经在做智能导

购和智能内容方向的技术原型及论证，智能导购在技术上主要是借鉴对话系统，通过引导用户和引擎对话与关键字推荐方式互为补充，包括自然语言理解，对话策略，

对话生成，知识推理、知识问答和商品搜索等模块，功能主要包括：

a. 根据用户搜索上下文生成引导用户主动交互的文本，比如搜索“奶粉”时，会生成“您宝宝多大？0~6 个月，6 个月到 1 岁……”引导文案，提示用户细化搜索意图，如果用户输入“3 个月”后，会召回相应段位的奶粉，并在后续的搜索中会记住对话状态“3 个月”宝宝和提示用户“以下是适合 3 个月宝宝的奶粉”。

b. 知识导购，包含提高售前知识问答或知识提示，比如“3 个月宝宝吃什么奶粉”回答“1 段”。目前对话技术正在提高中，尤其是在多轮对话状态跟踪、知识问答和自动评价几个方面，但随着深度学习、强化学习和生成对抗学习等技术在 NLP、对话策略、阅读理解等领域的应用，越来越多的训练数据和应用场景，domain specific 的对话技术未来几年应该会突飞猛进。智能内容生成，包括生成或辅助人工生成商品和清单的“卖点”，短标题和文本摘要等，让淘宝商品表达更加个性化和多元化。语义搜索。语义搜索主要是解决关键字和商品内容之间的语义鸿沟，比如搜索“2~3 周岁宝宝外套”，如果按照关键字匹配召回结果会远小于实际语义匹配的商品。

语义搜索的范围主要包括：

a. query tagging 和改写，比如新品，年龄，尺码，店铺名，属性，类目等搜索意图识别和归一化，query tagging 模型是用的经典序列标注模型 bi-lstm + CRF，而标签分类（归一化）作为模型另外一个任务，将序列标注和分类融合在一起学习。

b. query 改写，主要是计算 query 之间相似度，把一个 query 改写成多个语义相似的 query，通常做法是先用不同改写策略生成改写候选 query 集合，比如词替换、向量化后 top k、点击商品相似度等，然后在用 ltr 对后续集合排序找出合适的改写集合，模型设计和训练相对简单，比较难的

是如何构建高质量的训练样本集合，线上我们用 bandit 的方法探测部分 query 改写结果的优劣，离线则用规则和生成对抗网络生成一批质量较高的样本。

c. 商品内容理解和语义标签，通过商品图片，详情页，评价和同义词，上下位词等给商品打标签或扩充商品索引内容，比如用 image tagging 技术生成图片的文本标签丰富商品内容，或者更进一步用直接用图片向量和文本向量融合，实现富媒体的检索和查询。

d. 语义匹配，经典的 DSSM 模型技术把 query 和商品变成向量，用向量内积表达语义相似度，在问答或阅读理解中大量用到多层 LSTM + attention 做语义匹配，同样高质量样本，特别是高质量负样本很大程度上决定了模型的质量，我们没有采样效率很低的随机负采样，而是基于电商知识图谱，通过生成字面相似但不相关的 query 及相关文档的方法生成负样本。从上面可以看到 query tagging、query 相似度、语义匹配和语义相关性是多个目标不同但关联程度非常高的任务。下一步计划用统一的语义计算框架支持不同的语义计算任务，具体包括：

a. 开发基于商品内容的商品表征学习框架，为商品内容理解，内容生成，商品召回和相关性提供统一的商品表征学习框架，重点包括商品标题，属性，详情页和评价等文本信息抽取，图像特征抽取和多模信号融合。

b. query 表征学习框架，为 query 类目预测，query 改写，query 推荐等提供统一的表征学习框架，重点通过多个 query 相似任务训练统一的 query 表征学习模型。

c. 语义召回，语义相关性等业务应用模型框架。语义搜索除了增加搜索结果相关性，提升用户体验外，也可以一定程度上遏制淘宝商品标题堆砌热门关键词的问题。

智能匹配。这里主要是指个性化和排序。内容包括：

a. ibrain (深度用户感知网络)，搜索或推荐中个性化的特点是用户的理解与表达，基于淘宝的用户画像静态特征和用户行为动态特征，我们基

于 multi-modals learning、multi-task representation learning 以及 LSTM 的相关技术，从海量用户行为日志中直接学习用户的通用表达，该学习方法善于“总结经验”、“触类旁通”，使得到的用户表达更基础且更全面，能够直接用于用户行为识别、偏好预估、个性化召回、个性化排序等任务，在搜索、推荐和广告等个性化业务中有广泛的应用场景，感知网络超过 10B 个参数，已经学习了几千亿次的用户行为，并且会保持不间断的增量学习越来越聪明。

b. 多模学习，淘宝商品有文本、图像、标签、id、品牌、类目、店铺及统计特征，这些特征彼此有一定程度的冗余和互补，我们利用多模学习通过多模联合学习方法把多维度特征融合在一起形成统一的商品标准，并在多模联合学习中引入 self-attention 实现特征维度在不同场景下的差异，比如女装下图片特征比较重要，3C 下文本比较重要等。

c. deepfm，相对 wide & deep 模型，deepfm 增加了特征组合能力，基于先验知识的组合特征能够应用到深度学习模型中，提升模型预测精度。

d. 在线深度排序模型，由于行为类型和商品重要性差异，每个样本学习权重不同，通过样本池对大权重样本重复 copy 分批学习，有效的提升了模型学习稳定性，同时通过融合用户状态深度 ltr 模型实现了千人千面的排序模型学习。

e. 全局排序，ltr 只对单个文档打分然后按照 ltr 分数和打散规则排序，容易导致搜索结果同质化，影响总页效率，全局排序通过已知排序结果做为上下文预测下一个位置的商品点击概率，有效提升了总页排序效率。

f. 另外工程还实现了基于用户和商品向量的向量召回引擎，相对倒排索引，向量化召回泛化能力更强，对语义搜索和提高个性化匹配深度是非常有价值的。以上实现了搜索从召回、排序特征、排序模型、个性化和重排的深度学习升级，在双 11 无线商品搜索中带来超过 10% (AB-Test) 的搜索指标提升。

多智能体协同学习实现智能决策

搜索中个性化产品都是成交最大化，导致的问题是搜索结果趋同，浪费曝光，今年做的一个重要工作是利用多智能体协同学习技术，实现了搜索多个异构场景间的环境感知、场景通信、单独决策和联合学习，实现联合收益最大化，而不是此消彼长，在今年双 11 中联合优化版本带来的店铺内和无线搜索综合指标提升 12% (AB-Test)，比非联合优化版本高 3% (AB-Test)。

性能优化。在深度学习刚起步的时候，我们意识到深度模型 inference 性能会是一个瓶颈，所以在这方面做了大量的调研和实验，包括模型压缩(剪枝)，低秩分解，量化和二值网络。通过以上技术，今年双 11 期间在手淘默认搜索、店铺内搜索、店铺搜索等均取得了 10% (AB-Test) 以上的搜索指标提升。

阿里巴巴人工智能搜索应用的未来计划

通用用户表征学习。前面介绍的 DUPN 是一个非常不错的用户表征学习模型，但基于 query 的 attention 只适合搜索，同时缺少基于日志来源的 attention，难以推广到其他业务，在思考做一个能够适合多个业务场景的用户表征模型，非搜索业务做些简单 fine tuning 就能取得比较好的效果；同时用户购物偏好受季节和周期等影响，时间跨度非常大，最近 K 个行为序列假设太简单，我们在思考能够做 life-long learning 的模型，能够学习用户过去几年的行为序列；

搜索链路联合优化。从用户进入搜索到离开搜索链路中的整体优化，比如 搜索前的 query 引导（底纹），搜索中的商品和内容排序，搜索后的 query 推荐（锦囊）等场景；

跨场景联合优化。今年搜索内部主搜索和店铺内搜索联合优化取得了很好的结果，未来希望能够拓展在更多大流量场景，提高手淘的整体购物体验；多目标联合优化。搜索除了成交外，还需要承担卖家多样性，流量

公平性，流量商业化等居多平台和卖家的诉求，搜索产品中除了商品搜索外还有“穹顶”，“主题搜索”，“锦囊”，“内容搜索”等非商品搜索内容，不同搜索目标和不同内容（物种）之间的联合优化未来很值得深挖。



2017 年的语音识别，路只走了一半

作者 | Tina

这一年，百度开放了语音平台DuerOS，阿里补贴了4个亿销售百万智能音箱抢占语音入口。而作为语音识别的先驱龙头，大家开始担心科大讯飞用近二十年建立起来的技术壁垒被摧毁，有人扒讯飞的业务，有人开始扒讯飞十年的财报……

这一年的人工智能大潮，无疑让更多人关注科大讯飞，关心在这样的潮流里，一家深耕语音识别的公司如何能获得更多业务和利润，如何能去迎合AI上升的趋势，从而满足人们对人工智能的所有期望。

其实这一年，技术的进程还是和往年一样。（我们从语音识别的角度来解读2017年的进展，部分技术解读来源自对讯飞的采访）。

2017，从数据提升开始说起

去年IBM、微软、谷歌和百度都发布过自家语音识别进展，而今年对

媒体更新词错率进展的有三家：

2017年3月，IBM结合了LSTM模型和带有3个强声学模型的WaveNet语言模型。“集中扩展深度学习应用技术终于取得了5.5%词错率的突破”。相对应的是去年5月的6.9%。

2017年8月，微软发布新的里程碑，通过改进微软语音识别系统中基于神经网络的听觉和语言模型，在去年基础上降低了大约12%的出错率，词错率为5.1%，声称超过专业速记员。相对应的是去年10月的5.9%，声称超过人类。

2017年12月，谷歌发布全新端到端语音识别系统（State-of-the-art Speech Recognition With Sequence-to-Sequence Models），词错率降低至5.6%。相对于强大的传统系统有16%的性能提升。

大家的目标很一致，就是想“超过人类”，之前设定人类词错率为5.9%的这个界线。

总结来说，因为Deep CNN引入之后，语音识别取得了很大的突破，例如谷歌从2013年到现在，性能提升了20%。

而国内语音识别的企业如百度、搜狗、科大讯飞，识别率都在97%左右。在语音识别这件事情上，汉语比英语早一年超越人类水平。

去年，科大讯飞又推出了全新的深度全序列卷积神经网络（DFCNN）语音识别框架，该框架的表现比学术界和工业界最好的双向RNN语音识别系统识别率提升了15%以上。今年，在实际应用领域，讯飞输入法的识别准确率在今年7月份也终于突破了97%，达到了98%。

技术“可用”是第一步，但技术最终是要落地的，变成产品和服务才能实现价值。

今年技术应用场景有什么变化？

今年的产品落地，让人联想到的首先肯定是智能音箱。

2016年的数据统计表明，中国智能音箱销售量占全球比重为0.35%，6万：1710万台的差距。在2017年双十一阿里的补贴销售之后，终于可以

说“中国智能音箱销量在百万以上”，“中国的智能音箱得到了爆炸式的增长”。但从需求上说，智能音箱的功能集中在听音乐、闹钟、智能家居等，这些功能并不属于国人的“刚需”。BAT巨头都将智能音箱作为语音入口进行抢占，也给了我们一种爆发的假象。

但这一年，应用场景无疑是越来越丰富。基于各个领域的应用拓展，智能语音技术已经走出安静的室内或者私人环境，走上了服务大厅、卖场及行驶中的汽车等。技术的应用也越来越深入。机器翻译、远场识别、智能降噪、多轮交互、智能打断等技术的进步，也给智能语音的应用场景带来了更多的变化。

在智能车载领域，2017年科大讯飞发布的飞鱼系统2.0，融合了 Bargain全双工语音交互技术，窄波束定向识别技术，自然语义理解技术，免唤醒技术，多轮对话技术等科大讯飞核心技术。目前，科大讯飞已经为超过200款车型，累计超过1000万部车辆输出了语音交互产品。

此外，在新零售领域，智能语音技术的应用也在不断扩展。比如12月18日，科大讯飞和红星美凯龙发布战略合作计划，未来由科大讯飞研发的智能导购机器人“美美”将在全国红星美凯龙门店上市。

语音识别六十年，技术突破总是艰难而缓慢

语音识别的研究起源可以追溯到上世纪50年代，AT&T贝尔实验室的Audry系统率先实现了十个英文数字识别。

从上世纪60年代开始，CMU的Reddy开始进行连续语音识别的开创性工作。但是这期间进展缓慢，以至于贝尔实验室的约翰·皮尔斯（John Pierce）认为语音识别是几乎不可能实现的事情。

上世纪70年代，计算机性能的提升，以及模式识别基础研究的发展，促进了语音识别的发展。IBM、贝尔实验室相继推出了实时的PC端孤立词识别系统。

上世纪80年代是语音识别快速发展的时期，引入了隐马尔科夫模型（HMM）。此时语音识别开始从孤立词识别系统向大词汇量连续语音识

别系统发展。

上世纪90年代是语音识别基本成熟的时期，但是识别效果离实用化还相差甚远，语音识别的研究陷入了瓶颈。

关键突破起始于2006年。这一年辛顿（Hinton）提出深度置信网络（DBN），促使了深度神经网络（Deep Neural Network, DNN）研究的复苏，掀起了深度学习的热潮。2009年，辛顿以及他的学生默罕默德（D. Mohamed）将深度神经网络应用于语音的声学建模，在小词汇量连续语音识别数据库TIMIT上获得成功。2011年，微软研究院俞栋、邓力等发表深度神经网络在语音识别上的应用文章，在大词汇量连续语音识别任务上获得突破。国内外巨头大力开展语音识别研究。

科大讯飞的智能语音探索之路

科大讯飞在2010年首批开展DNN语音识别研究，2011年上线了全球首个中文语音识别DNN系统。2012年，在语音合成领域首创RBM技术。2013年又在语种识别领域首创BN-ivec技术。2014年科大讯飞开始深度布局NLP领域，2015年，RNN语音识别系统全面升级。

2016年，上线DFCNN（深度全序列卷积神经网络，Deep Fully Convolutional Neural Network）语音识别系统。在和其他多个技术点结合后，科大讯飞DFCNN的语音识别框架在内部数千小时的中文语音短信听写任务上，相比目前业界最好的语音识别框架双向RNN-CTC系统获得了15%的性能提升，同时结合科大讯飞的HPC平台和多GPU并行加速技术，训练速度也优于传统的双向RNN-CTC系统。DFCNN的提出开辟了语音识别的一片新天地，后续基于DFCNN框架，还将展开更多相关的工作。

DFCNN的结构如图所示，它直接将一句语音转化成一张语谱图作为输入，即先对每帧语音进行傅里叶变换，再将时间和频率作为图像的两个维度，然后通过非常多的卷积层和池化（pooling）层的组合，对整句语音进行建模，输出单元直接与最终的识别结果比如音节或者汉字相对应。



图 1

在语音识别子领域上，今年科大讯飞的智能语音技术所取得的代表性的成就在自然语言理解领域。7月份，哈工大讯飞实验室(HFL)刷新了斯坦福大学发起的SQuAD(Stanford Question Answering Dataset)机器阅读理解挑战赛全球纪录，提交的“基于交互式层叠注意力模型”(Interactive Attention-over-Attention Model)取得了精确匹配77.845%和模糊匹配85.297%的成绩，位列世界第一，也是中国本土研究机构首次取得赛事榜首。

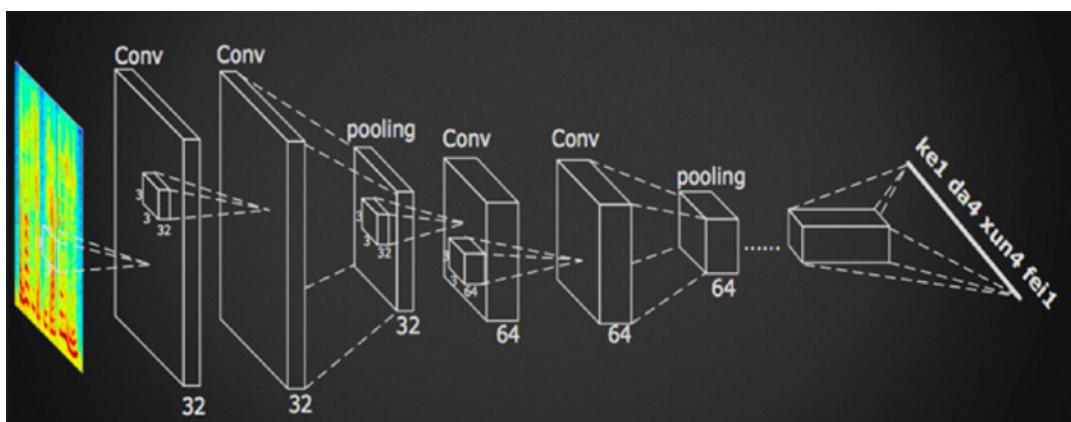


图 2

语音合成上，暴风雪竞赛(Blizzard Challenge)是国际最权威的语音

合成比赛。科大讯飞以语音合成技术率先达到4.0分的成绩并连续12年蝉联全球第一名，这是全世界唯一能让语音合成技术能够达到真人说话水平的系统。5.0分代表播音员的水平，4.0分代表美国普通老百姓的发音水平。

在人机交互系统上，科大讯飞于11月发布了AIUI2.0系统，支持远场降噪、方言识别和多轮对话的技术的基础上又增加了主动式对话、多模态交互、自适应、个性化识别等能力并能在嘈杂会场完成全双工翻译功能。

而科大讯飞的云端语音开放平台，截至2017年12月，累计终端数达到15亿，日均交互次数达到40亿，开发者团队数已达50万。

语音识别还有哪些没有解决的问题？

深度学习应用到语音识别领域之后，词错率有显著降低，但是并不代表解决了语音识别的所有问题。认识这些问题，想办法去解决，是语音识别能够取得进步的关键所在，将 ASR（自动语音识别）从“大部分时间仅适用于一部分人”发展到“在任何时候适用于任何人”。

1. 口音和噪声

语音识别中最明显的一个缺陷就是对口音和背景噪声的处理。最直接的原因是大部分的训练数据都是高信噪比、带有口音的语言。比如单是为美式口音英语构建一个高质量的语音识别器就需要 5000 小时以上的转录音频，因而仅凭训练数据很难解决掉这个问题。

在中国，口音问题解决得比较好的，是科大讯飞。科大讯飞目前推出了22种方言相关的语音识别系统，但对于那些音素体系与汉语不同的方言或外国语种，在成本问题上还没有很好的办法。

2. 多人会话

每个说话人使用独立的麦克风进行录音，在同一段音频流中不存在多个说话人的语音重叠，这种情况下的语音识别任务比较容易。然而，人类即使在多个说话人同时说话的时候也能够理解说话内容。一个好的会话语

音识别器必须能够根据谁在说话对音频进行划分（Diarisation），还应该理解多个说话人语音重叠的音频（声源分离）。

在利用语音技术推动输入和交互模式变革的过程中，仍面临这些阻碍。多人对话等场景下的语音识别率虽然很高，声纹识别虽然也已经在实验室实现，但距离实际应用还有一些距离。

3.认知智能

语音识别技术在质检、安全等方面有很好的应用，但是对于人类所希望达到100%的识别率来说，从科研角度看肯定还有很多需要继续努力的地方。比如减少语义错误、理解上下文上（机器的学习和推理），我们才仅触及皮毛。“认知智能有没有真正的突破，是这一轮人工智能热潮——包括产业化热潮——能不能进一步打开天花板、进一步形成更大规模的产业的关键技术所在”，2017年底，科技部正式发文将依托科大讯飞建立首个认知智能国家重点实验室。

未来五年内，语音识别领域仍然存在许多开放性和挑战性的问题，如，在新地区、口音、远场和低信噪比语音方面的能力扩展；在识别过程中引入更多的上下文；Diarisation 和声源分离；评价语音识别的语义错误率和创新方法；超低延迟和高效推理等。尽管语音识别目前成果斐然，但剩下的难题和已克服的一样令人生畏。虽然近几年深度神经网络的兴起使得语音识别性能获得了极大的提升，但是我们并不能迷信于现有的技术，总有一天新技术的提出会替代现有的技术。

除技术外，一个 AI 企业的那些事儿

人工智能催生了大量新技术、新企业和新业态，人工智能火热背景下，作为A股人工智能龙头股科大讯飞，曾一个月猛增360多亿元，市值突破千亿。似乎很正契合普通百姓对“AI”神化的认知。

2017年11月15日，中国新一代人工智能发展规划暨重大科技项目启动会在京召开，科技部公布我国第一批国家人工智能开放创新平台，包括：

1. 依托百度公司建设自动驾驶国家新一代人工智能开放创新平台；
2. 依托阿里云公司建设城市大脑国家新一代人工智能开放创新平台；
3. 依托腾讯公司建设医疗影像国家新一代人工智能开放创新平台；
4. 依托科大讯飞公司建设智能语音国家新一代人工智能开放创新平台。作为首批入选国家新一代人工智能开放创新平台，目前的科大讯飞，用刘庆峰的话说是“现在还未到达登顶的状态，只能说是已经开始登山，刚克服了爬坡之后的艰难，开始到慢慢适应的状态”，如同语音识别技术现状。

人工智能是个大趋势，本身也是需要很重投入的，但它也会有更长远的影响，所以不能特别短视于此时此刻的回报上。“必须具备了强技术，才能形成刚需”，“就是要把技术做深做透，做到大家真正觉得有刚需”，刘庆峰说，“我们瞄准着五到十年更前沿的技术研究”。



2017 年，吴恩达都干了些啥？我们替他做了份年终报告

作者 | Debra

2017年过去了，人工智能四大金刚之一吴恩达（Andrew Ng）这一年可谓风风火火，3月份出走百度曾引起轩然大波，一度掀起是不是AI风口要来的讨论，也有不少人认为他离开百度是为了加盟其妻创办的Drive.ai。然而，吴恩达离开之后陆续创办了DeepLearning.ai和Landing.ai，三大人工智能项目已经实现了两个。过去的一年中，吴恩达都忙了些什么呢？我们来替他做了一份年终总结。

离开百度之前

吴恩达在今年3月份离开百度之前，手头完成了一些重要的工作，如在CES大会上推出了对话式人工智能操作系统DuerOS，并与合作伙伴小鱼在家共同推出第一款搭载该系统的智能家庭服务人。这与吴恩达让AI惠

及每一个人的理想不谋而合。



AndrewNg吴恩达 V

1月5日 14:35 来自 微博 weibo.com

很高兴在CES2017推出百度对话式人工智能操作系统DuerOS并且与我们的合作伙伴小鱼在家共同推出第一款搭载DuerOS系统的智能家庭机器人。2017年毫无疑问将是对话式操作系统迅猛发展的一年。我们期待通过这种合作让人工智能技术进入更多的家庭，惠及每一个人。



1月份，他带领的项目“百度大脑”机器人在综艺节目《最强大脑》中与最聪明的人类进行人机大战，最后以三回合两胜一平的成绩胜出。虽然是一档综艺节目，难免有些哗众取宠的味道，但是百度大脑在人机大战中胜出，还是一定程度上证明了百度AI的实力。



AndrewNg吴恩达 V

1月21日 11:55 来自 微博 weibo.com

三场人机大战落幕 @百度大脑 三回合两胜一平。最后一场弱光下动态识别项目无论对王昱珩还是小度来说都很困难，没想到小度正确识别出了两次，期待#百度大脑#未来在刑侦方面可以发挥实质性的功能。感谢@最强大脑 这个节目和所有参赛选手，你们一直在刷新我们对人脑的认知，由衷的钦佩你们的超强脑力。



除了机器人，吴恩达似乎对移动AR非常看好，1月17日，他发布推特

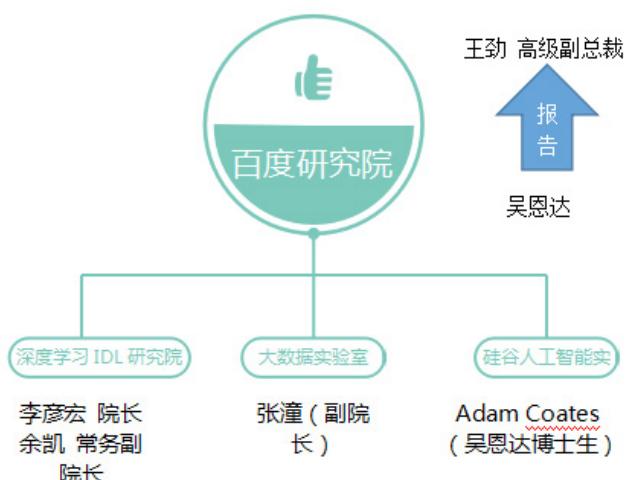
宣布成立AR Lab，而且在多篇推文中表达了他对于移动AR，尤其是中国移动AR前景的乐观态度。

吴恩达为什么放弃百度，要自己另立门户呢？熟悉他的人可能都知道他在担任百度首席科学家这两年，可能是因为自己想要在百度落实自己学术上的研究没有得到理想的效果，他决定离开。吴恩达在百度的人事关系是怎样的呢？我们来做一下简单的总结。

总结他在百度的经历，要从百度研究院建立之初开始说起，那时吴还没有加入百度。

2013年，余凯（后离开百度，创立地平线）带领百度成立深度学习IDL研究院。在余凯的引荐下，吴接受了百度的offer，正式加盟百度，担任公司的首席科学家兼研究院院长，负责百度研究院的领导工作，向高级副总裁王劲汇报。

吴恩达加入百度之前，百度研究院主要在北京，最开始仅有一个深度学习主题的研究院，即IDL研究院，李彦宏亲任院长，余凯系常务副院长。之后，张潼加入，又以副院长的身份组建了大数据实验室。再之后，跟着吴恩达2014年5月一起加入百度的斯坦福博士生Adam Coates，又在硅谷组建了硅谷人工智能实验室。



由此，百度研究院演变成在北京的深度学习研究院IDL、大数据实验

室和硅谷人工智能实验室三个分支。

那么问题来了，硅谷和北京怎么分工？如何合作？出现冲突如何做决定？尽管吴恩达名义上时首席科学家和研究院院长，但余凯、张潼等研究院副院长级别高，在工业界的历练更多，也更年长，在各自领域也有自己的专长和班底，这都是事实。至于百度研究院内部的人事关系到底如何，吴是否感觉受到牵制，为何出走的真实原因，我们就不得而知了。

离开百度后

Deeplearning.ai——教育，人才培养

3月份，吴恩达离开百度。在离开百度之际，吴恩达在微博上发表了文章《开启我在人工智能领域的新篇章》，总结百度在人工智能领域的成就，他说，他的团队在过去的两年中，每年都孵化出一项新业务：一项是无人驾驶，另一项是DuerOS语音交互计算平台。

当时，人们纷纷猜测他此后的去处，大多数人认为他可能会转向其妻 Carol Reiley（卡罗尔·莱利）创办的自动驾驶软件创业公司Drive.ai。



但他们只猜对了一半。6月，吴恩达加入Drive.ai任董事会成员。Drive.ai由斯坦福大学的多名人工智能研究员创立于2015年，基于深度学

习的技术研发驱动无人驾驶的自动驾驶软件系统，并在2016年4月获得了加州车管局的自动驾驶测试许可。这家创业公司创办者Carol Reiley无需多加介绍，她的生平已经被扒过无数遍了，生于美国，祖籍台湾省高雄市，研究生在霍普金斯大学做医疗机器人研究，用AR把触觉信号变为视觉信号，辅助医生来判断操作医疗机械时的力量大小，后转投自动驾驶软件研究开发，是一位名副其实的美女+女神级的人物。

但吴并没有把大部分的精力放在这家公司，而是专注于开创自己的AI天地。

 **AndrewNg** 吴恩达 
8月9日 09:08 来自 微博 weibo.com

今天deeplearning.ai正式启动，在Coursera平台上推出了深度学习系列课程。我希望用deep learning赋能众多人工智能爱好者，构建AI驱动的社会。



 收藏  321  95  662

6月，吴通过Twitter、微博等平台正式发布了Deeplearning.ai，揭开了他三大人工智能项目中第一个项目的神秘面纱。这是一个专注于AI教育和AI人才培养的平台，由五门课组成深度学习系列课程，旨在推广普及深度学习知识，并通过他创建的Coursera网站向大众提供最新在线课程。当时吴恩达说，Deeplearning.ai的使命，是向全世界普及深度学习知识。现在，国内对深度学习感兴趣的人也可以通过网易云课堂观看教学课程，目前在该平台可以观看5门课程。

官网链接：<https://www.coursera.org/specializations/deep-learning>

网易云课堂链接：https://mooc.study.163.com/university/deeplearning_ai#/c

目前来看，Deeplearning.ai项目在Coursera推出，使得这个项目得以平稳发展，并迎来了一些合作伙伴。比如英伟达，其在11月宣布与

Deeplearning.ai建立了新的合作关系，双方将共同研发新的机器翻译培训材料，以教育人工智能人才，解决AI人才紧缺的现状。

这是他对于AI惠及每一个人愿景的实践。国家的根基在于教育，AI的发展同样离不开人才的培养，从长远来说，Deeplearning.ai可以说是承载他AI愿景最具有战略意义的行动，功在千秋。

AI Fund, L.P.——投资，资金支持

根据美国证券交易委员会官网信息显示，8月14日，吴恩达注册了一支专注投资人工智能的风投基金 AI Fund, L.P.，规模达 1.5 亿美元，从资金层面支持AI的发展。在AI发展势头正劲的当头，他用这支基金布下了他的AI格局中的一颗棋子。

Capital Markets
Company Overview of AI Fund, L.P.

December 25, 2017 3:52 AM ET

Snapshot	People																		
<p>Company Overview AI Fund, L.P. is a fund of AI Fund, L.P.</p>	<p>Key Executives For AI Fund, L.P. Dr. Andrew Y. Ng Investment Professional</p>																		
Compensation as of Fiscal Year 2017. United States																			
<p>Similar Private Companies By Industry</p> <table border="1"> <thead> <tr> <th>Company Name</th> <th>Region</th> </tr> </thead> <tbody> <tr> <td>@Visory LLC</td> <td>United States</td> </tr> <tr> <td>1 Road Partners LLC</td> <td>United States</td> </tr> <tr> <td>11T Partners, LLC</td> <td>United States</td> </tr> <tr> <td>123Jump.com, Inc.</td> <td>United States</td> </tr> <tr> <td>1509225 Ontario, Inc.</td> <td>United States</td> </tr> </tbody> </table>	Company Name	Region	@Visory LLC	United States	1 Road Partners LLC	United States	11T Partners, LLC	United States	123Jump.com, Inc.	United States	1509225 Ontario, Inc.	United States	<p>Recent Private Companies Transactions</p> <table border="1"> <thead> <tr> <th>Type</th> <th>Date</th> <th>Target</th> </tr> </thead> <tbody> <tr> <td colspan="3">No transactions available in the past 12 months.</td> </tr> </tbody> </table>	Type	Date	Target	No transactions available in the past 12 months.		
Company Name	Region																		
@Visory LLC	United States																		
1 Road Partners LLC	United States																		
11T Partners, LLC	United States																		
123Jump.com, Inc.	United States																		
1509225 Ontario, Inc.	United States																		
Type	Date	Target																	
No transactions available in the past 12 months.																			

据彭博社关于企业的信息页面显示，这支基金在最近12个月没有任何交易，可惜该网站并没有这支基金董事会和委员会的详细信息。

另据相关人士了解，吴恩达对 AI Fund, L.P. 描述的模式是：每个点子都是他的，他会作为co-founder拿一定的股份（至少40%），另一位co-founder也会拿一些股份，资金由基金来出，按照标准（明星项目）A轮估

值。

比如说Landing.ai可能估值2亿美金，基金出1000万占股5%，富士康出1000万占股5%，吴自己出任CEO，占股50%以上。另外他作为基金GP还要分一部分钱。这个项目找到团队并启动之后，他就再去创办下一个公司。所以，他可以利用这个基金投资自己感兴趣的项目，并占据大部分股份。由此看来，吴可并不是一个科技型书呆子。

Landing.ai——AI与制造业碰撞出火花



AndrewNg 吴恩达 V
 12月14日 22:03 来自 秒拍网页版
 我很高兴地宣布Landing.ai的成立。作为一家人工智能（AI）公司，Landing.ai旨在帮助企业在人工智能时代实现转型。我们从制造业开始！#人工智能##数字化制造# □ AndrewNg吴恩达的秒拍视频



12月14日，吴恩达又扔下一枚重磅炸弹，宣布成立Landing.ai，同时宣布其已经与富士康在7月份达成了战略合作，将集两家的核心能力打造AI技术、人才和系统，包括产业链的管理、工人的培训等多个方面。

Landing.ai是迄今为止吴恩达完成的第二大人工智能项目，而这一次，他把关注点放在了让AI技术在制造业实现落地、解决失业率的问题上。在微博和Twitter上发布Landing.ai的前几分钟，他曾发表了一篇名为《通过人工智能复兴制造业》的文章，在文中，他解释了他为何选择AI与制造业结合的原因，以及对AI在传统行业中的应用现状和未来进行描述。



目前，根据Landing.ai官网发布的团队介绍照片，这个项目除了创始人兼CEO吴恩达之外，已经吸引了不少成员加入，而且还在继续招兵买马，放出的招聘职位包括机器学习、视觉、全栈等软件工程师，自动化和机器人方向的机械工程师，以及培训业务经理等。

2017年，吴恩达离开百度后，也许是为了证明在百度落实不了的梦想，他在一年之内完成了他所说的三大项目中的其中两个，也不禁让人好奇他的第三个项目要在什么时候揭开庐山真面目。虽然不大可能在今年全部完成，但12月的小尾巴还有一周，不到最后一刻谁知道人工智能领域还会发生什么惊喜呢？



2017 开发者盘点：在他们眼中，AI 是吹牛灌水还是真的来了？

作者 | 蔡芳芳 陈利鑫

写在前面

2017 年可以说是 AI 元年，几乎所有的科技大事件都发生在 AI 领域。

还记得当初你被一个会后空翻的机器人支配的兴奋吗？

后来，OpenAI 的研究人员又突发奇想，让机器人相互摔跤、踢足球、翻墙，据说效果非常不错。这颤抖的小腿、风骚的走位是不是很萌？

会出汗冷却系统的机器人，话说你真的需要健身减肥吗？

以及不久前获得了沙特阿拉伯居民身份的索菲亚，她会开“我要毁灭人类！”的玩笑话（惊出一身冷汗），还会暗戳戳嘲讽机器威胁论的忠实拥护者马斯克。

自动驾驶技术也真正走到了聚光灯下，赚足了眼球！

事实上，这些比较具有舆论效应的全民热议话题只是冰山一角，AI

领域在今年发生的变革隐藏在嘈杂噱头之下，等待我们去挖掘总结。

为此，我们对 AI 领域的数十位开发工作者进行了采访，问了问他们 AI 领域的哪些变化在他们眼里可以被视为里程碑式的事件？对于有人诸如 AI 论文 99% 是灌水，99% 媒体是吹牛的言论，它们是怎么看的等问题，来看看他们怎么说的吧！

印象篇

「各大公司都在宣称自己在 AI 方向的加注，以及民众，媒体对'AI'的狂欢。」

2017 年，AI 领域给开发者人员留下最深刻印象的有哪些事件呢？

在各位开发人员的回复中，我们发现几乎每个人对 2017 年 AI 领域印象最深刻的事件都包括 AlphaGo 和 AlphaGo Zero 战胜人类围棋选手这一历史性事件。对此，也有人表示，“今年印象最深刻的莫过于 AlphaGo 了，虽然觉得其实结果没太大意义，但带来的影响确实大。”

另外，AI 程序打败德州扑克专家、人脸识别、波士顿 Atlas 机器人后空翻、智能音箱、量子计算机、百度发布无人驾驶平台 Apollo 进行无人驾驶车辆测试、中国将人工智能上升为国家战略等事件也都给他们留下了深刻的印象。

也许一位微软的语音工程师的回答最能从广泛意义上反映出 AI 在国内的现状：“最让我印象深刻的，是各大公司都在宣称自己在 AI 方向的加注，以及民众，媒体对'AI'的狂欢。”

这是 AI 开发者对于 2017 年 AI 领域发生事件的印象，那么对于普通大众来说，这一年他们对于 AI 的印象是什么样的呢？

可能普通大众对于 AI 的印象还是很高大上的东西，“民众对 AI 的理解主要来自于科幻电影，而多数媒体为了生存和点击率，需要写民众爱看的内容，那么从主观上就有吹牛的动机，从而失去了新闻工作者客观的职业素养。所以现在网上充斥着“震惊！xxxx；重磅！xxx；恐怖！xxxx”这种文章。不可否认的是，这些文章的确提供了一些有用的信息，比如会提供一

些有意思的工作的链接。不过通常过于浮夸的报道，会加剧民众对 AI 不切实际的想象，以及影响一些经验不足的从业人员，比如学生。”微软的一位语音识别工程师说道。

炒作篇

「技术炒作这种事情，如果一点都没有，也不见得就是好事。」

AI 在近几年的炒作很厉害这一点无须否认，在这一领域工作多年的技术人员见证了 AI 是怎么被炒火起来的过程。在他们的印象中，AI 领域都经历过那些炒作呢？

移动电商 算法工程师 —— 图像算法技术研发

我参加工作 11 年了，有印象的技术炒作包括：智能家居、3D 打印、机器学习、人脸识别、VR、无人驾驶、虚拟试衣等。有的技术经过长时间的沉淀找到了应用场景，比如机器学习用于推荐、人脸识别用于金融。有的技术估计还是概念，没有落地。总的来看，炒作普及了大众的技术视野，尽管有泡沫，泡沫过后总会留下一些东西，或者是技术的、或者是商业的，这些会长期支持技术商业化。

IBM 前端工程师 —— 云计算，AI，Social，前端 +.Net

我在工作中见过的炒作包括前端框架、Vs studio 各种推陈出新等。对此，我的感悟是底层很重要，不管什么框架，重点还是要会看原生的代码。技术平均每年大更新一次，找准自己的方向，尽快转型。看到做 mainframe 的同事眼睁睁看着自己的技术被淘汰，太迟转型面临失业降薪，很心痛，要吸取教训。

Coupang 数据工程师 —— 电商

我工作已经超过 10 年，经历过的 AI 炒作有 Java EE、开源、大型分布式电商平台、大数据、机器学习 / 深度学习、AI。我的感受是技术发展太快，需要学习的东西太多，有时候为了照顾广度，就很难有深度；有些领域要深入需要一定的时间和实践，如果太专注某个领域，又担心技术过时，赶不上潮流。

BAT 研发工程师 —— 搜索 +AI DevOps、敏捷实施和落地

工作六年多，经历过云计算、容器、DevOps、AI 等技术理念的兴起到成熟。

每年我都会关注 Gartner 的 Hyper Cycle，有点可惜的是，Gartner 的 Hyper Cycle 对中国不太热衷，我们只能借鉴美国和印度市场的趋势了。从这几年看，中国 IT 的发展越来越贴近美国发展的趋势。特别是从容器、DevOps、AI 的兴起来看，我们真的没有落后太多了。所以 Gartner 的 Hyper Cycle 差不多就是一张能告诉我们这波码农，未来我们需要掌握什么领域或者往那里跳槽的风向标了。

而其他的技术，例如区块链技术，还是要想好跟自己的工作能结合，才会去学习。

还有很多技术人员见证的炒作还包括数据挖掘、区块链、量子计算、AlphaGo、无人驾驶、物联网、AR/VR 等。

“感觉虽不靠谱，但是得有人无畏前行，人多自然能出结果”，“虽然各大公司都在努力的追求技术的创新，但是如何把新的技术在实际生产中应用，是目前应该解决的问题”，“技术炒作这种事情，如果一点都没有，也不见得就是好事”，是这些技术人员在历经无数热潮时内心最真实的感受。

与 AI 结合技术门槛篇

「有一个名词称为奥卡姆剃刀，说的是假如有多个模型都能达到差不多的效果，选最简单的。」

对于那些结合了 AI 的公司，他们是如何在工作中应用 AI 的？结合 AI 的过程有难度吗？是否有技术门槛呢？

明略数据 AI 工程师 —— 文本挖掘

我在工作中做自然语言处理，总体来说，处理无结构的文本数据还是很有难度的，就算 AI 技术，也无法像一个成年人一样，迅速理解各种灵活的表达方式和新词。

蘑菇街 算法工程师 —— 电商公司，面向年轻女性提供时尚化和社交化的购物平台

我目前负责图像算法技术的研发工作，工作中主要是结合图像识别技术来改善电商平台的产品体验。AI 有难度，也有技术门槛，不仅仅是指技术本身，而是指 AI 落地的门槛，要对应用场景有足够的理解，有业务的敏锐感觉。

百度 AI 工程师 —— 搜索 手机百度 app 和简单搜索 app 的图像客户端

本人负责百度移动端的 AI、图像等工作，与 AI 紧密相关。2017 年以前移动和 AI 结合难度较大，瓶颈主要在于移动设备的性能较差无法承担起较大的神经网络运行。随着移动设备性能不断增强，各大公司跟进开发框架。在软硬件双飞的未来，移动设备和 AI 结合已是大势所趋。

腾讯 后台开发 —— 手机浏览器游戏、app 个性化分发

我负责 QQ 移动浏览器内的游戏、app 的个性化推荐，需要使用到较多的机器学习的算法，和传统后台相比，AI 的门槛还是比较高，尤其对非科班出身的同学，需要补齐大量的高数、概率统计、优化理论等数学基础，还需要了解统计学习、深度学习等常用的算法。

用一句话概括结合 AI 的技术门槛：“有的放矢，否则不仅毫无意义，而且难度很大”，也有人认为“传统的技术和目前大热的 AI 技术相结合，有一定的难度，掌握一些计算机基础知识的人更容易入门 AI 技术。”

但也有人表示，“工作中与 AI 的结合不难，但是要做到准确率高、高并发响应很难”，“关键在于数据，技术难度不算太大”，但毕竟这种说法是少数的。

总之，大多数技术人员认为 AI 与目前的工作结合起来难度较大，技术门槛较高，而且难点在于与实际应用场景的结合。而在机器学习界，有一个名词称为奥卡姆剃刀，说的是假如有多个模型都能达到差不多的效果，选最简单的。而用高端模型不一定能充分理解业务，反而是采用简单模型的效果更好。因此，企业在实际应用场景中应该根据自己的业务类

型，选择适合自己的模型。

人才篇

现在，关于 AI 人才稀缺，院校毕业生刚出校门就可以年薪百万千万的消息刺激着不少人的神经，对于 AI 培训课程趋之若鹜，面对高昂的培训费用也非常舍得砸钱进修。那么，AI 人才的缺口到底有多大？这些开发人员怎么看待目前市场上 AI 人才的超高薪资和各类 AI 入门培训课程呢？

在我们的采访对象中，不少人表示自己并不是出身 AI 专业的，也没有 AI 相关学历，他们入门 AI 的方法主要是通过在实践工作中积累经验，加上观看网络视频教学课程、阅读相关源代码、技术博客等方式自学 AI，有着自学不息的精神。

对于十分火爆的 AI 培训课程，其中一些回答描述非常与现实状况实际上十分贴切，“近年来一些大数据的课程带上 AI 的帽子会变得更加抢手，但实际上教授的还是大数据相关的东西。但课程本身如何还是基于个人需求。”

而这些课程是否真的对于入门 AI 有帮助呢？大部分回答持否定的态度，认为两三个星期的培训就可以从事门槛相对较高的 AI 工作有些不现实，AI 薪资虚高，课程质量也参差不齐。

明略 数据工程师

“AI 入门培训都是在骗钱，收智商税，想学的话踏踏实实补数学，踏踏实实看书看博客，精读几篇经典的 paper，github 上多复现几个实验。”

AWS 工程师 —— 云计算

“泡沫太严重了，AI 还没有实际创造那么大价值。”

也有人表示离自己太遥远，无法做出评价。

但也有人对 AI 课程火爆的现象给予了比较客观的评价。

第四范式 工程师 —— 机器学习平台

AI 人才的高薪说明行业对这类稀缺人才的关注，实际上在任何行业

顶尖的人才都是稀缺的，也会有所谓的超高温的情况，这点我认为是正常的并且对我个人没有太大影响。AI 培训也是适应这个行业潮流所逐渐出现的，有意思的是以前宣言是大数据培训的课程现近一般都会加上 AI 或者人工智能这个招牌，但实际上还是培训大数据相关的技术，不同人对这类课程的需求也是不一样的，我个人更推荐通过参与开源设计以及实践的方式来入门机器学习领域。

“这么火，谁都会来分一杯羹的。就像当初大数据炒作一样，超高温是因为人才太少。入门课程还是很有用的，但是要有干货。”

也有人在校有过两年相关科研经历的硕士在读生提醒在校生不要盲从，要根据实际情况和自己的需要选择学习的方式。“目前 AI 技术是一个发展前景非常好的产业，各大公司不惜高薪来吸引 AI 人才的加入，但是随着时间的推移，AI 人才的积累势必会饱和，所以在校的大学生等也要有一些清晰的意识，不要盲从。如果说自己可以学习的东西就不需要培训了，当然如果自学很吃力，可以考虑报辅导班。”

局限篇

「在真实的业务场景中，要落地机器学习模型的门槛还是比较高。」

当问及结合各自目前的工作中，AI 目前存在的最大局限是什么时，这些开发人员的回答中不少都涉及到了 AI 落地难的问题，当然也有人提到了 AI 过于依赖于深度学习，方法单一、缺乏训练数据、拓展性差、识别伪 AI 等问题。

明略数据 工程师

AI 在工作中最大的局限在于落地，真正落地赚钱的产品屈指可数，这个行业健康发展，还是要多接地气，切实的为社会解决问题。”

第四范式 工程师

在真实的业务场景中，要落地机器学习模型的门槛还是比较高，用户需要能够把业务模型抽象成机器学习任务，并且完成数据接入、数据清洗、特征抽取、模型训练、参数调优等一系列过程，对于业务理解能力和

开发能力都有很高的要求，而且除了模型训练其他目前都难以使用 AI 来协助解决。因此我认为未来 AI 的发展应该向着低门槛的方向，通过算法或者工程的优化，例如实现易用的数据引入和处理平台、实现 AutoML 和自动调参的算法，让更多人可以真正把 AI 落地实际业务中。

AWS 工程师

太依赖于深度学习了，方法单一，有些问题深度学习效果也不是很好，估计未来几年学术界会出现比深度学习效果更好的方法。

浪潮软件 数据采集工程师

对无规则事物的分析学习。毕竟图像、围棋、声音都还是有一定之规的，仍需要人脑将抽取规则输入给 AI。

蘑菇街 算法工程师

目前所谓 AI，都是基于人类经验或者积累的数据来工作的，只能做简单的重复的工作，下围棋也不例外，只是在人类能力上的复制。未来的 AI，应该是具有一定推理能力的，例如智能助理。

此外，腾讯的开发后台工程师则表达了对监管法规空缺的忧虑。

影响篇

「AI 的出现让很多开发者心浮气躁，但脚踏实地的人不多。」

AI 的火爆发展在开发人员眼里是什么样的？他们如何看待 AI？AI 对于普通开发者的工作，是正面影响多还是负面影响多？综合采访对象的回答，我们得出结论，大多数人认为 AI 对他们的积极影响还是大于消极影响的。

“AI 是人们的合作伙伴，不可能取代人类。普通开发者，应该了解 AI，利用 AI 改善自己的工作，比如开发工具、开发技能，以及解决问题的思路。AI 不能取代开发者，但是可以辅助开发者。”——明略数据 工程师

“我还不知道，我觉得我现在做的工作（编程）将来可能会被 AI 取代。但是一定有新方向可以供我发展，虽然我还没看到这个方向在哪

里。”——IBM 前端工程师

但也有人表示，AI 的出现让很多开发者心浮气躁，但脚踏实地的人不多。

“我觉得 AI 是一种技术，就像前端后端安卓 IOS 一样，不希望大家跟风学，现在这个行业很浮躁，真正的脚踏实地的人很少，大部分都在仰望星空，这不是好现象，所以对于普通的开发人员来说，负面影响多，但是我还是看好这个行业。”——明略工程师 2

误解篇

「人们认为 AI 可以上天入地，无所不能，仿佛很快就会有一支 AI 军队出现一样。」

说起普通大众对 AI 最大的误解，可能就是认为 AI 比人更智慧，无所不能，认为机器和真人一样会思考，超越人类，控制人类，取代人类的工作等。

就如 IBM 一位前端工程师所说：“人们认为 AI 可以上天入地，无所不能，仿佛很快就会有一支 AI 军队出现一样。实际上，AI 是要 train 的！要 train 的！要大量的 training 的！重要的话说 3 次。AlphaGo 级别的 AI，需要有一个团队加上很多高级计算机在 train，AI 不是随随便便找条流水线就能量产的。”

“人们认为人工智能是来取代人的，会伤害人，这是很大的误解。技术本身无对错，这取决于怎么使用 AI，谁来使用 AI。AI，也是人类智慧的产物。只是如果被坏人利用了的话，AI 会伤害人类的利益；比如隐私的泄露和滥用、甚至有 AI 武器的出现。”——蘑菇街算法工程师

“目前大家对于 AI 技术感到新奇并带有一点恐惧，主要因为不了解底层的原理和细节，误以为是“自我觉醒”的模型。AlphaGo 并不是自学掌握围棋规则的，也不会自己悄悄地用电脑下棋，其实它是一颗蒙特卡洛树加上一个精心设计的神经网络模型，而这个程序只是接受棋盘信息作为输入，然后输出赢棋概率最大的落子点而已，本质上 AlphaGo 只是一些列

浮点数而已。对于 AI 技术原理不了解，并且加上个人臆造的想象，是目前非 AI 从业者普遍的最大误解。”——第四范式 工程师

最期待应用篇

「他们最期待的 AI 应用中无人驾驶的呼声最高。」

对于这些专业的技术开发人员，他们对于 AI 应用的期待也是五花八门，多少也能代表一部分人的心声。

未来，他们最期待的 AI 应用中无人驾驶的呼声最高，很多受访者不约而同看好无人驾驶的发展，甚至有人打趣道为了等待无人驾驶的到来至今没有考驾照！也有不喜欢做家务的人表示希望拥有保姆、管家类 AI，可以分担家务；也有不少人对 AI 音箱、自动编码 AI、个人助理等表示很感兴趣。

学习计划篇

「多关注机器学习和深度学习，将之与目前的工作结合起来。」

作为专业的开发人员，面临层出不穷的新技术和工作要求，他们也有意识地在加强 AI 方面的学习，自我学习对其发展的重要性也越来越重要。那么，在 2018 年，他们有什么学习计划呢？

结合各自的工作领域，他们分别在学习实变函数、泛函分析、量子力学、汇编、计算机视觉、js 全栈、最优化理论 +TensorFlow、机器学习基础理论及 AI 在金融和医疗领域的应用等方面有学习的计划。

总之，大多数技术人员认为多关注机器学习和深度学习，将之与目前的工作结合起来，是 2018 年他们需要做的事情。

最后，对于 AI 的宣传，大多数采访对象认为媒体有过分夸大之嫌，甚至有人认为 AI 的 paper 99% 是灌水，AI 的媒体 99% 是跟风，AI 的团队 99% 存在着落地难，刷排名现象，背后有资本推动的原因。

对于媒体上部分宣传，可能是行业外人士因为信息不对称产生的误解，比如之前误报的 Facebook 机器人发明新语言的事，其实不过是模型

训练的时候跑飞了，普通的自动化工厂被宣传为人工智能无人工厂，而一些根据统计信息生成文本的程序也被捧为具有自我意识的人工智能机器人，这些都存在过分夸大的成分。

但是，也有人认为目前 AI 已经真实来临，这个事实不存在虚假或者吹牛成份。

AI 到底是灌水吹牛的成分多，还是会在 2018 年真切地来到我们的身边呢？我们不妨静静等待，看 AI 在 2018 年将会带来什么惊喜！



扫码关注InfoQ公众号

Geekbang | **InfoQ**
极客邦科技