

中国AI基础数据服务行业发展 报告

2020年



1

目前人工智能商业化在算力、算法和技术方面基本达到阶段性成熟，想要更加落地，解决行业具体痛点，需要大量经过标注处理的相关数据做算法训练支撑，可以说数据决定了AI的落地程度。

2

对比中国与世界的发展情形来看，人工智能行业发展前景良好，而作为强关联性的AI基础数据服务行业受其发展红利的影响，未来市场仍有不小的上升空间。

3

2019年中国AI基础数据服务行业市场规模可达30.9亿元，预计2025年市场规模将突破100亿元，年化增长率为21.8%。

4

中小型数据供应商的整体体量仍然可观，但随着业务门槛提升、客户需求多样化，越来越多的“中小型数据供应商”在苦恼生存问题，该群体在未来1-2年内将迎来一阵“倒闭潮”。

5

“品牌数据服务商”在这1-2年内应该“居安思危”，注重品牌认可度塑造、提高规模化生产能力、大力发展如预标注、项目进度可视化等精细化运营方式、增加如语音合成（TTS）数据处理等差异化服务，最终追求的是单位价格的利润最大化。

中国及全球人工智能发展概览

1

人工智能基础数据服务市场现状

2

人工智能基础数据服务行业壁垒

3

行业发展机遇与挑战

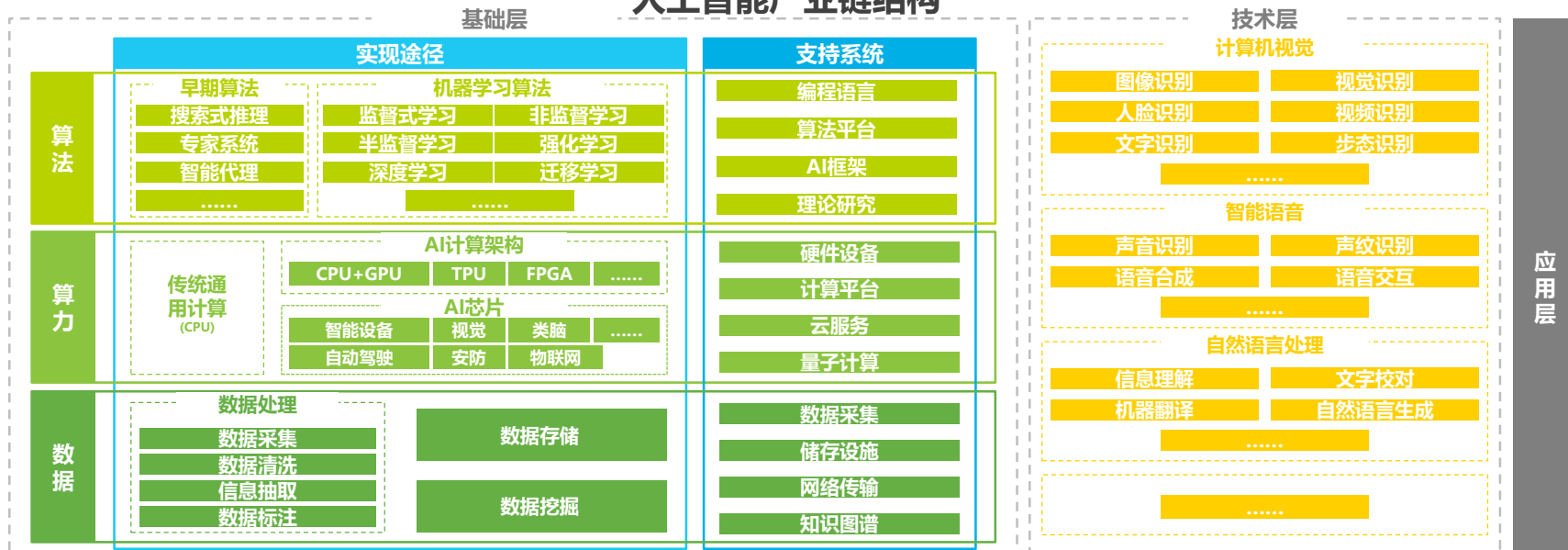
4

人工智能产业发展和产业链结构

数据决定了AI落地程度，是商业化过程中重要的一环

自从2012年深度学习在图像和语音方面产生重大突破后，人工智能便真正具备了走出实验室步入市场的能力，2016年AlphaGo的胜利再次引爆行业，成功唤起了中国市场的兴趣，时至今日，人工智能的商业化在中国得到了长足发展，在安防、金融、企服等领域纷纷落地开花，同时也真正意义上衍生出了一套完整的产业链。AI产业链可以分为基础层、技术层和应用层，基础层按照算力、数据和算法再次划分，对整体上层建筑起到支撑作用；技术层根据算法用途分为计算机视觉、智能语音、自然语言处理等，是AI最引人注目的环节；应用层则按照不同场景的需求定制开发专属服务，是AI真正赋能行业的方式。目前人工智能商业化在算力、算法和技术方面基本达到阶段性成熟，想要更加落地，解决行业具体痛点，需要大量经过标注处理的相关数据做算法训练支撑，可以说数据决定了AI的落地程度，而AI基础数据服务行业又鲜有关注，因此本报告承接艾瑞《2019年中国人工智能基础数据服务白皮书》，再次挖掘该行业的现状和发展，展示其真实的一面。

人工智能产业链结构



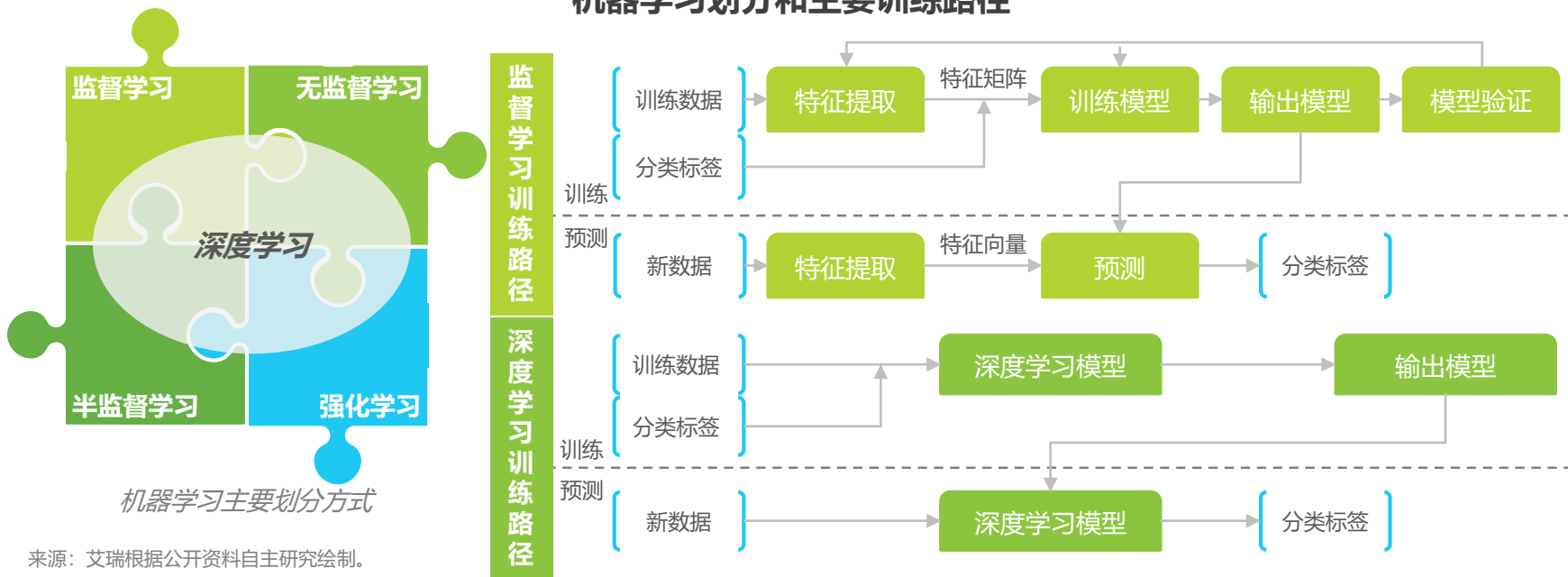
来源：艾瑞根据公开资料自主研究绘制。

人工智能技术实现路径

机器学习是主流，其中监督学习下的深度学习是主要方式

人工智能是对一类能够实现机器模拟智慧生命某些特征的技术统称，从学术上可以分为以知识工程为代表的符号主义、以神经网络为代表的连接主义和以仿生机器人代表的行为主义三个流派，近些年掀起新一轮人工智能热潮的机器学习就属于连接主义学派。机器学习按照训练方式可分为使用人工标注分类标签训练的监督学习、无分类标签且自动聚类推断的无监督学习、使用少量人工标注+自动聚类的半监督学习和根据现实情况自动“试错+调整”的强化学习四类，而最著名的深度学习同样是机器学习的分支，但因为模型结构的不同而与上述训练方式不在一个区分范畴，深度学习在训练方式上均可与四种方式发生重叠。目前来看AI应用最广泛的计算机视觉和智能语音更依赖于监督学习下的深度学习方式，半监督和无监督是学术界尝试突破的方向，当下仅在如无人驾驶中急转弯场景训练等特定领域中得以尝试应用，而强化学习被认为是更接近人类在自然界中学习知识的方式，在最佳路径选择、最优解探寻等方面有所应用，但泛化能力还有待突破。

机器学习划分和主要训练路径



来源：艾瑞根据公开资料自主研究绘制。

人工智能对数据资源服务的需求趋势

定制化需求成为主流，数据服务市场步入需求常态化

监督学习下的深度学习算法训练十分依赖人工标注数据，2012年-2016年期间人工智能行业不断优化算法增加深度神经网络层级，利用大量的数据集训练提高算法精准性，ImageNet开源的1400多万张训练图片和1000余种分类在其中起到重要作用，为了继续提高精准度，保持算法优越性，市场中产生了大量的标注数据需求，这也催生了AI基础数据服务行业的诞生。时至今日，人工智能从业公司的算法模型经过多年的打磨，基本达到阶段性成熟，随着AI行业商业化发展，更具有前瞻性的数据集产品和高定制化数据服务需求成为了主流。据了解，目前一个新研发的计算机视觉算法需要上万张到数十万张不等的标注图片训练，新功能的开发需要近万张图片训练，而定期优化算法也有上千张图片的需求，一个用于智慧城市的算法应用，每年都有数十万张图片的稳定需求；语音方面，头部公司累计应用的标注数据集已达百万小时以上，每年需求仍以20%-30%的增速上升，不仅如此，随着IoT设备的普及，语音交互场景越来越丰富，每年都有更多的新增场景和新需求方出现，对于标注数据的需求也是逐步增长。结合市场来看，随着AI商业化发展，AI基础数据服务需求步入常态化，存量市场具有较为稳定的需求源头，而增量市场随着应用场景的丰富，以及新型算法的诞生，拥有更广阔的想象空间。

人工智能对数据资源服务的需求趋势

需求变化

人工智能算法模型基本达到阶段性成熟，随着AI落地行业的趋势，更具有前瞻性的数据集产品和高定制化数据服务需求成为了主流

前瞻性数据集产品
定制服务

存量市场

结合计算机视觉和语音市场的需求情况看，AI基础数据服务需求成为常态，主要需求方稳定，数据需求量保持稳步增长

稳步增长

增量市场

随着AI落地传统行业，以及IoT设备的普及，人工智能的应用场景越来越丰富，需求方类型也在逐年增加，具有广阔的想象空间

空间广阔

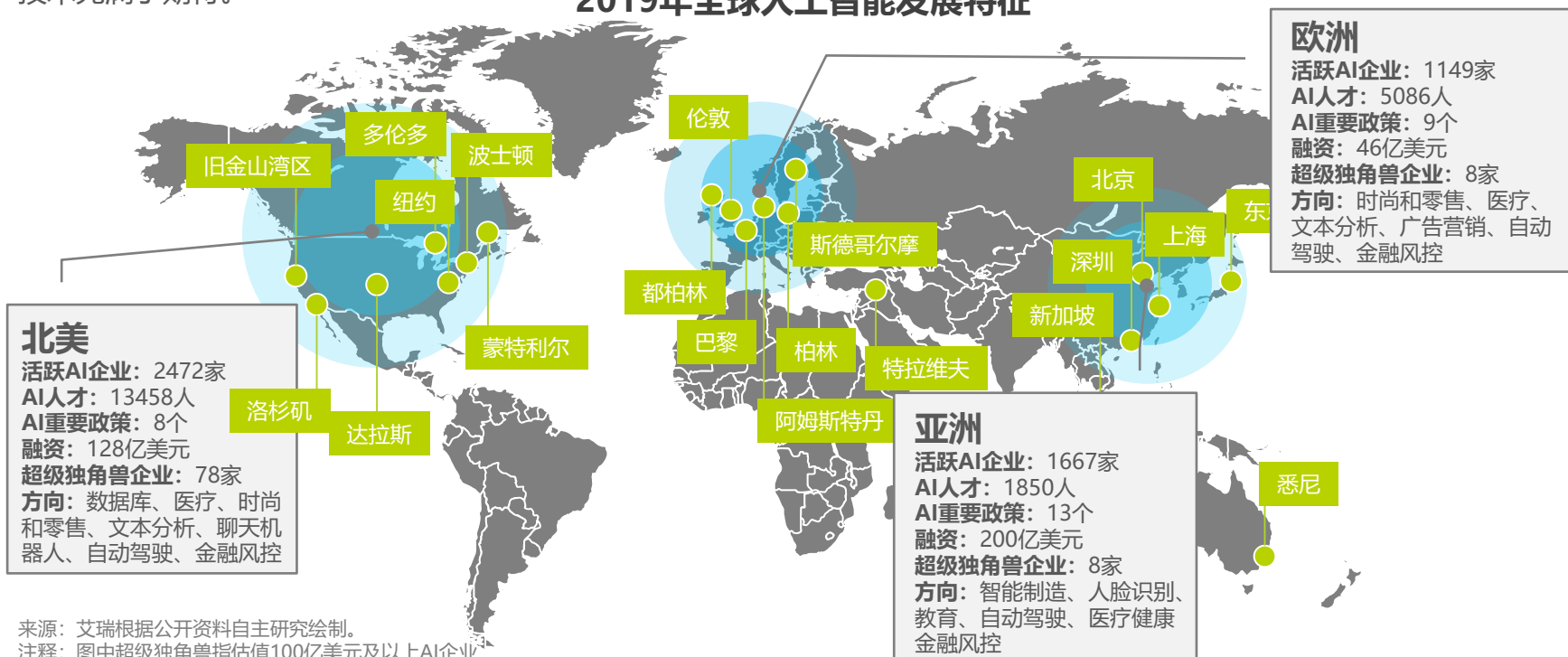
来源：艾瑞根据公开资料自主研究绘制。

全球人工智能产业发展情况

全球市场火热依旧，呈现集群式发展，AI落地仍是主旋律

2019年全球人工智能行业发展依旧火热，重点围绕北美、欧洲和亚洲三大区域发展，共有5386家活跃企业，27400名高级研究人员，20座重点发展城市，产生了约374亿美元融资，其中自动驾驶、药物医疗、人脸识别、视频内容和金融反欺诈是获得融资最多的领域。综合来看，美国仍然是人工智能领域的领头羊，每年不仅向全球输出最前沿的AI应用概念，提供最接近市场需求的应用场景，同时也吸收着世界各地不断涌入的人才、数据资源和创意灵感。以英国、德国为代表的欧洲发展区和以中国、日本、新加坡、印度为代表的亚洲发展区整体实力相近，但走出了两条完全不同的路径，虽然欧洲储备了大量的AI人才，但市场活性不强，对于创新业务的发展并不友好，而亚洲虽然人才储备不充裕，但市场活力十足，对AI技术充满了期待。

2019年全球人工智能发展特征



来源：艾瑞根据公开资料自主研究绘制。
注释：图中超级独角兽指估值100亿美元及以上AI企业

中国与全球人工智能发展情况对比

中国AI发展属世界前列，并仍处于上升期，前景良好

全球人工智能发展了六十年，中国参与了二十余年，随着政府意志和市场意志双重聚焦，中国的人工智能发展进入了黄金期。对比2019年中国与全球人工智能发展情况，在AI相关论文发布数量、企业数量、融资总额、产业规模、专利申请数量等方面中国均居世界头部阵营，具有充分的市机场竞争力。补足短板方面，中国也在不断努力。人才储备是技术软实力发展的核心，教育问题并非能一蹴而就地解决，目前全国已经有35所高等院校开设了AI专业，国际交流和国际人才引进也在不断加深，未来5年内将有大量从业者涌入市场。产业结构单一也是中国人工智能行业存在的固有问题，研发型企业远少于应用型企业的隐患随着中美专利竞争而浮现，政府开始重视AI基础层创业公司的培养，资本方也更加关注AI芯片、机器学习算法、数据处理等产业链上游企业的发展，科技巨头企业更是提前进行了AI生态布局，建立了产业联盟，在各方的努力中中国AI市场处于从局部向整体发展的上升期，行业前景良好。

2019年中国与全球人工智能发展情况对比



论文数/FWCI引用均数 占全球AI论文28%/FWCI为0.8

人才数/高级人才数 博士413名/专家120名

活跃企业数 1189家

融资总额 约166亿美元

AI论文占全部论文的3%

论文数

博士22000名/专家5400名

人才数/高级人才数

5386家

活跃企业数

约374亿美元

融资总额



预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_20953

