

人工智能产业链分析报告

一、什么是人工智能

人工智能从1956年达特茅斯会议中诞生，几经起伏。2016年3月，AlphaGo计算机程序轻取围棋九段棋手李世石，立刻引发全世界的讨论。让机器能展示与人类相似的认知的AI能够驾驶汽车，也会盗取隐私；能推动企业的生产力也能加强企业的侦查能力。人工智能可以将工人从重复的或者危险的劳动中解放出来，将极大推动生产力的提高，但同时也激发了对人工智能或将取代人类工作的焦虑情绪，甚至有人担心人类最终会创造出连自己都无法控制的智能机器。在纷繁的观点背后，有一点毋庸置疑：人工智能有着改变全球社会的巨大潜力。

人工智能（Artificial Intelligence）是指使用机器代替人类实现认知、识别、分析、决策等功能，其本质是对人的意识与思维的信息过程的模拟，是一门综合了计算机科学、生理学、哲学的交叉学科。凡是使用机器代替人类实现认知、识别、分析、决策等功能，均可认为使用了人工智能技术。

随着数据资源增长、计算能力的提升、算法的优化，目前人工智能已经进入部分领域应用阶段。比如 Apple 的 Siri，谷歌的无人车，IBM 的 Watson，以及其它各种人脸识别技术等。作为一种基础技术，理论上讲人工智能能够被应用在各个基础行业（如 AI+金融、AI+医疗、AI+传统制造业等），同时也有其如机器人这样具体应用行业的概念。人工智能正催生新的业态和商业模式，引发产业结构的深刻变革，成为新一轮工业革命的推动器。

二、人工智能历史沿革——逾 60 年历史，再次步入黄金发展期

人工智能发展迄今已有 60 余年历史，由于发展瓶颈逐渐突破，2013 年起进入黄金发展期。

第一次黄金发展期：人工智能诞生于 1956 年的达特茅斯会议，该会议同时诞生了人工智能最初的成就和最早期的研究者。1956 年~1972 年经历了人工智能发展的第一次黄金时期。这一阶段机器学习、神经网络和人工智能领域得到探索与突破，研究者获得大量经费支持。1957 年罗森布拉特发明出第一款神经网络 Perceptron，将人工智能推向第一个高峰。

第一次低谷：1972 年人工智能遭遇第一次低谷，研究者们遭遇许多问题——计算机计算能力不足，难以解决任何实际的 AI 问题；数据库难以满足 AI 应用

需求；现有 AI 逻辑框架无法解决常见问题等。这一系列问题导致政府及资助机构对人工智能失去信心，停止相关领域研究补助。

第二次黄金发展期：1981 年~1987 年，人工智能迎来第二次繁荣，专家系统和知识工程在全世界迅速发展，为企业等用户赢得巨大的经济效益，1986 年 BP 神经网络算法由 Rumelhart 和 McClelland 为首的科学家小组提出，该算法使得大规模神经网络的训练成为可能，将人工智能推向第二个黄金期。

第二次低谷：20 世纪 80 年代末，各国争相进行的智能计算机研究计划先后遇到严峻的挑战和困难，AI 又一次遭遇财政问题，这促使人工智能研究者们对已有的人工智能思想和方法进行反思，这样的反思有助于人工智能迎来第三次高潮。

第三次黄金发展期：20 世纪 90 年代以来，人工智能在各子领域悄然发展，1997 年 IBM 制造的电脑“深蓝”击败了国际象棋冠军，2006 年 Hinton 首次提出“深度学习”神经网络，使得人工智能性能获得突破性进展。从目前来看，深度学习是实现人工智能最有效、取得成效最大的实施方法。深度学习提出后，大数据云计算等基础技术也不断进步，迎来了 2013 年以来的人工智能第三次高潮。

三、三大驱动要素相继突破瓶颈，人工智能已至爆发节点

（一）算法：模型性能更优，巨头开源意在“集思广益”

1. 机器学习模型的建立过程

开发机器学习模型的过程可分为两个阶段：训练阶段（Training）和推理阶段（Inference）。

训练阶段（Training）又可以细分为 3 个步骤——第一步开发人员需要定义一组函数，这一过程实际上就是在组建神经网络模型；第二步开发人员把海量数据所组成的数据集（文字、图形、视频等各种形式的的数据）输入到模型中，然后经过计算输出最终结果后将输出结果与预定的目标结果进行对比，这一过程也就是在训练上一步建好的模型，初步建成的模型所输出的结果和预定要达到的目标结果之间存在着误差，在第二步最后要定义一个总误差函数；第三步，计算并调整模型中的函数参数，使得第二步模型输出值与目标值之间的总误差函数值最小，使得总误差最小的参数集合将被选定为“最佳”函数 f^* 的参数，参数选定后“最佳”函数 f^* 也随之确定。

完成训练阶段的这3个步骤以后，还需要用训练数据集之外的其他数据测试“最佳”函数 f^* 在遇到从未接触过的数据时其输出结果是否依然准确。如果输出结果是准确的，则说明训练阶段所得到的“最佳”函数 f^* 是一个“学习”能力比较强的模型；如果输出结果不准确，说明机器只能识别自己“见”过的数据而不能识别出没见过数据（即“过拟合现象”），机器还没有具备较好的学习能力，需要开发人员再返回修改模型及参数，直至测试结果准确为止。这也就是上面所提到的第二阶段——“推理阶段”（Inference）。

2. 算法和层级是模型性能更优

从1957年第一个神经网络——感知器被提出以来，便不断有新的算法涌现，尤其是在上世纪80年代，以及20世纪初至今，新算法出现频率比较高。优质算法的不断迭代使得模型的准确率不断提高，另一方面神经网络自身隐含层数的增加也明显提升了模型性能及模型对现实的刻画能力。优质算法的不断迭代可以使得模型的准确率得到明显提升，例如：2010年深度学习网络（DNN）替代了GMM-HMM模型，使得语音识别的词错误率明显下降。基于GMM-HMM（高斯混合模型-隐马尔可夫模型）的算法的语音识别词错误率一直维持在21%以上居高不下，但是自从2010年神经网络算法的进一步研究，词错误率大幅下降至14%-19%的水平（虽然人类听写的词错误率只有2%-4%）。百度从2012年开始也不断将深度神经网络（DNN）、长短期记忆模型（LSTM）、深度卷积神经网络（CNN）等新算法应用至语音识别技术中，使得除了算法迭代，算法自身隐含层数的增加也可以提高模型准确率。在2015年底，微软亚洲研究院的研究员提出了一个新的基于CNN的深度模型——深度残差网络，将深度学习神经网络的隐含层数从几十层推进到152层，一举夺得当年ImageNet 图像识别大赛的第一名，模型的错误率（3.57%）大约为第二名（6.7%）的一半。

然而构建算法模型需要耗费巨大的时间和精力，而近几年各大知名互联网科技公司以及科研机构都相继开源了自己的深度学习框架，大大缩短了构建算法的时间。像谷歌Tensorflow、伯克利Caffe、微软CNTK 等等这样的深度学习框架实际上是一个神经网络模型库，里面汇集了各种各样的神经网络模型，而每一个神经网络模型就好比是一套完整的积木，这套积木里的每一个小木块就相当于是神经网络模型中的一部分，开发者可以从各套不同的积木里抽取不同的小木块然后

组建自己的完整积木模型。

（二）数据：数据“量”已足够，标注成本大成“隐忧”

数据和云存储遵从摩尔定律：每过两年，世界上的数据量就会翻倍，同时用于存储这些数据的成本则会以同样的速度下降。得益于互联网时代的到来以及移动互联网时代的发展，每天数据都已惊人的速度增长。IDC 报告显示，在 2013 年的人类产生数据是 4.4 个 ZB，到 2020 年这个数字将会增加到 44 个 ZB（1ZB=1012GB），复合年增长率达到 36%。现实生活中每 4 小时沃尔玛用户产生数据超过 2.5PB，每天 Twitter 产生推文有 5 亿条，今日头条的用户请求超过 60 亿次，每天今日头条处理数据量超过 6.3PB。实际上目前已经积累的数据量已经远远超过人工智能需求的数据量。

但是在目前人工智能算法还是以监督学习算法为主导的背景下，需要对数据进行标注才可用于模型训练，而且标记数据成本高昂，所以从这个意义上而言，虽然目前数据数量巨大并且还在高速增长，但是“可用”数据很少。监督学习需要带标记的数据集（在图像识别中，如果一张小猫的图片被研究人员标记为“猫”，并且输入到模型中用以训练模型，那么这里被标记为“猫”的图片就被称为带标记的数据）。

谷歌、Facebook、IBM 这样的科技巨头获取海量数据的成本很低，但是获取某些特定类别的带标记数据的成本依然无法显著降低。例如谷歌在 2017 年 3 月开放的语音数据库 AudioSet 里的语音片段全部来自于 YouTube（谷歌旗下子公司），获得这些数据几乎是无成本的；而 IBM 为了训练 Watson，不惜花重金收购了 Explorys 和 Phytel，从而获得了超过 1 亿份电子病历，可见医疗数据之“贵”。

训练机器学习算法模型所需要的数据量取决于要执行的任务、要满足的性能（要求模型所能达到的准确率）、所拥有的输入特征、训练数据中的噪音、提取特征中的噪音以及模型的复杂程度（例如神经网络的隐含层数、模型中参数的个数等）等因素。训练数据过少容易产生过拟合现象，所以要使模型达到一定的准确度，必须要使训练数据足够多。目前由于深度学习的兴起，神经网络的隐含层数大大增加，为使模型达到一定准确率，要求的训练数据也随之增加。2015 年微软亚洲研究院推出深度残差算法，将神经网络的层数由几十层推进到了 152 层，模型复杂度骤然增加，模型中神经元高达 2200 万个，连接有 113 亿（而 2012 年 8 层

神经网络只具有65万个神经元，连接为6亿）。

（三）计算力：GPU 为密集计算提速，专用芯片兴起大势所趋

训练深度学习算法的所需要的时间在很大程度上取决于硬件配置。训练一个模型大概需要1~3周的时间，训练一个复杂的模型的时间则更多（例如AlphaGo的策略神经网络和价值网络合计需要4周即一个月的时间）。而且在训练模型的过程中，往往需要尝试很多组参数，每调整一次就需要重新训练，所花费的时间将更长（例如在训练百度的机器翻译系统的过程中一共尝试了10组参数，那么整个训练时间就是100天）。如果计算资源不足以支撑大规模计算，那么对于产品更新换代和技术创新将十分不利。深度学习的测试阶段所需要的计算资源比训练阶段更多。例如AlphaGo测试阶段（比赛时）其单机版和人对弈时，需要用48块CPU和8块GPU，分布式版本则需要1200块CPU和176块GPU。

计算能力取决于计算机的硬件部分，而芯片是硬件体系的核心。目前用于机器学习的计算芯片主要有CPU（中央处理器）、GPU（图形处理器）、FPGA（现场可编程门阵列）、ASIC（应用型专用集成电路）四种架构。其中GPU是深度学习领域首选的芯片，在该领域拥有最高的市占率。目前，在芯片领域的技术发展在一定程度上提供了较为充足的算力。

四、政策密集发布，推动我国 AI 水平赶超国外

我国政府已经将人工智能的发展已经上升到了战略层面。从2015年5月份开始，国务院、发改委、工信部密集出台了一系列促进产业发展的政策。今年，人工智能产业发展受政府关注程度持续升温，2017年全国两会“人工智能”首次正式写入《政府工作报告》；2017年7月20日，国务院发布《新一代人工智能发展规划》，将新一代人工智能发展提高到国家战略层面，提出在2030年成为人工智能领域的世界领导者，核心产业规模超1万亿，带动相关产业规模超10万亿元；2017年7月23日，经国家科技部批准，人工智能产业技术创新战略联盟正式成立。

实施时间	颁布主体	法律法规	相关内容
2015.5	国务院	《中国制造2025》	组织研发智能制造装备以及智能化生产线，推动智能交通工具、智能工程机械、服务机器人、智

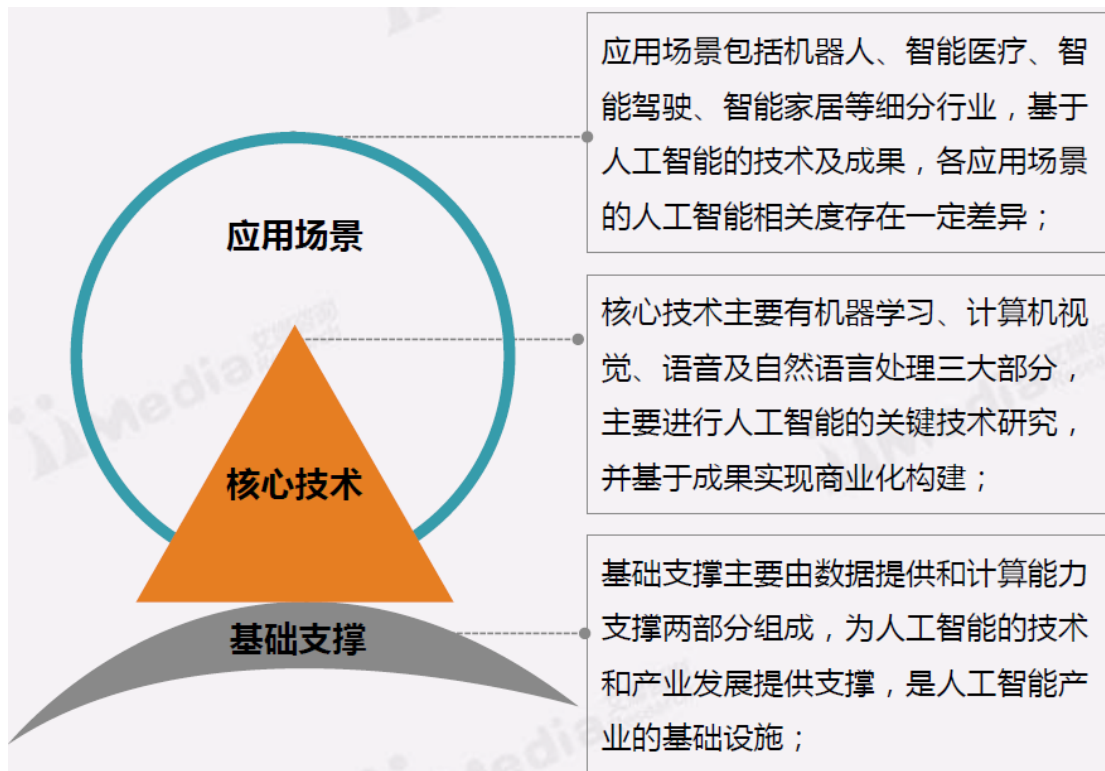
			能家电、智能照明电器、可穿戴设备等产品研发和产业化
2015. 7	国务院	《国务院关于积极推进“互联网+”行动的指导意见》	明确提出人工智能作为11个重点布局的领域之一，促进人工智能在智能家居、智能终端、智能汽车、机器人等领域的推广应用
2015. 7	中央办公厅、国务院	《关于加强社会治安防控体系建设的意见》	加大公共安全视频监控覆盖，将社会治安防控信息化纳入智慧城市建设总体规划，加深大数据、云计算和智能传感等新技术的应用
2016. 3	国务院	《国民经济和社会发展的第十三个五年规划纲要（草案）》	人工智能概念进入“十三五”重大工程，规划纲要提出要重点突破人工智能技术
2016. 4	工信部、发改委、财政部	《机器人产业发展规划（2016-2020）》	到2020年，自主品牌工业机器人年产量达到10万台。培育3家以上具有国际竞争力的龙头企业，打造5个以上机器人配套产业集群
2016. 5	发改委、科技部、工信部、网信办	《“互联网+”人工智能三年行动实施方案》	培育发展人工智能新兴产业，推进重点领域智能产品创新、提升中终端产品智能化水平，
2016. 7	国务院	《“十三五”国家科技创新规划》	研发智能机器人；发展大数据驱动的类人智能技术方法；开展下一代机器人技术、智能机器人学习与认知等技术研究
2016. 9	发改委	《关于组织申报	为构建“互联网+”领域创新网

		“互联网+”领域创新能力建设专项的通知》	络，促进人工智能技术的发展，应将人工智能技术纳入专项建设内容
2017.3	全国两会	《政府工作报告》	“人工智能”首次正式写入《政府工作报告》，报告中提出：“全面实施战略性新兴产业发展规划，加快新材料、新能源、人工智能、集成电路、生物制药、第五代移动通信等技术研发和转化，做大做强产业集群。”
2017.7	国务院	《新一代人工智能发展规划》	2020年，人工智能总体技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点 2025年，人工智能基础理论实现重大突破，部分技术与应用达到世界领先水平，人工智能成为我国产业升级和经济转型的主要动力 2030年，人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心

五、人工智能产业链的构成

人工智能的产业链结构可以分三个层次：基础支持层、核心技术层和应用场景层。基础支持层包括实现人工智能的数据资源、计算硬件以及算法模型，计算硬件主要由AI芯片、传感器等硬件构成。核心技术层则是为了实现机器完成对外部世界的探测，主要由计算机视觉、语音识别等感知层和语义识别等认知层构成，这些技术是机器能够做出分析判断的基础。应用场景层是集成了某种或多种基础应用技术的、面向如工业、自动驾驶、家居、仓储物流、金融、医疗等不同应用

场景的产品或方案。



六、核心技术不断突破，下游商业化进程加速

人工智能的核心技术应用包括语音识别、图像识别、语言理解和机器人等，深度学习等基础条件的持续突破使得核心技术取得重大进展，从而使商业化应用不断突破。语音识别和图像识别普遍正确率已超过 90%，目前正积极拓展下游商业化应用；语言理解是人机交互技术的基础，多元化的人机交互需求倒逼语言理解技术升级；机器人是工业 4.0 的支柱产业，政策推动下产业爆发在即。

（一）图像识别：技术不断成熟，人脸识别渐成焦点

根据科普中国撰写的对图像识别（计算机视觉）的定义，这是一门研究如何让机器“看”的科学，更进一步的说，是指用计算机代替人眼对目标进行识别、跟踪和测量的机器视觉，并进一步做图形处理，使计算机处理成为更适合人眼观察或传送给仪器检测的图像。图像识别应用场景可分为两大类：图像识别和人脸识别，每类又可继续划分为动、静共四个类别，基本覆盖了目前计算机视觉的各项应用场景。

图像识别技术不但有着非常高的识别准确率，而且能够很快给出智能的反馈，因此图像识别技术最容易快速落地到各行各业中。安防行业中的车辆数据提取，

医疗行业的影像诊断，电商行业中的精准营销，以及辅助驾驶都为图像识别技术提供许多落地变现的机会。人脸识别也是图像识别最重要的应用。人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术，用摄像机或摄像头采集含有人脸的图像或视频流，并自动在图像中检测和跟踪人脸，进而对检测到的人脸进行脸部的一系列相关技术。

（二）语音识别：应用前景广阔，寡头市场明显

语音识别是以语音为研究对象，通过信号处理和识别技术让机器自动识别和理解人类口述的语言后，将语音信号转换为相应的文本或命令的一门技术。语音识别其原理是：语音波形经过前端语音信号处理模块、声学模型模块、字典模块、语言模型四个关键的模块处理后，经过解码得到最终的语音识别结果。

DNN助力模型性能提升，但仍有多种困难待解决。从2011年开始深度学习技术的兴起使得语音识别的词错误率明显下降，性能不断提高。传统方法中用于特征提取的GMM模型逐渐被深度学习算法（DNN）所取代。2011年微软在声学模型的输出分布计算时，完全用神经网络的输出后验概率乘以一个先验概率来代替传统模型中的GMM的输出似然概率，使得词错误率大幅下降；自2011年微软开启了深度学习在语音识别中应用的先河之后，百度、谷歌等公司也纷纷上线用DNN进行声音建模的语音识别系统。目前语音识别在技术上仍面临诸多困难，主要包括噪声鲁棒性、多类复杂性、低数据资源、多语言特性、低计算资源等。

技术困难	描述
噪声鲁棒性	目前业内宣称的97%识别准确率，只在安静室内的近场识别中才能实现,但在一些噪声场景比较大的情况下，很难达到高识别准确率
多类复杂性	过去的大部分语音识别系统的设计主要是针对一些单一环境、单一场景下进行设计的，如何做多类别复杂场景下的通用的语音识别是非常困难的
多语言	目前大部分语音识别的研究和应用是基于一些大语种，但世界上共有6900多种语言，如何快速的实现一套基于任何语言的语音识别系统是非常困难
低计算资源	大部分的语音识别的应用都是基于云端的在线语音识别的服

务，基于有限的硬件资源做低功耗的离线的实时的连续性识别是非常困难的，

智能语音市场保持寡头垄断格局，市场集中度较高。从2002年语音识别技术应用刚刚起步到现在，全球语音识别市场竞争格局发生了巨大变化：（1）国外语音市场经历了行业并购整合：市占率在2002年排名第二的SpeechWork和排名第一的Nuance先后被ScanSoft收购，2005年ScanSoft更名为Nuance；（2）微软、谷歌、苹果等科技巨头公司随后纷纷布局语音识别，目前在语音识别市场上的地位已经显得举足轻重。在变化之中“不变”的是Nuance一直是全球语音识别市场的龙头公司。该领域市场集中度处于较高水平，就全球市场而言，排名前5的Nuance、谷歌、苹果、微软、科大讯飞所占市场份额达到80%以上，国内市场，排名前五的科大讯飞、百度、苹果、Nuance、捷通华声市场份额占比达到84.1%，其中科大讯飞以40%以上的国内市占率保持着绝对的龙头地位。

智能语音识别的应用场景丰富。智能家居可以通过语音控制一切家用设备；智能车载可以解决司机操作触摸屏过度分散注意力的问题；智能教育则可以进行口语测评和实时翻译；智能医疗可以为医生提供一个安全存储病历资料的云空间，方便语音搜索病历等。未来随着未来语音识别种类的进一步丰富，识别环境通用性的增强，以及远场语音技术的突破，一定会帮助拓展其应用范围到更多场景中。

（三）语言理解：人机交互需求倒逼技术升级，商业应用不断涌现

语言理解（自然语言处理）是研究人与计算机交互的语言问题的技术。与语音识别不同，语音识别是将语音信号转化为文本或相应指令，而语言理解是指让机器理解人类的语言，人们可以用自己最习惯的语言来使用计算机，无需再花费大量的时间和精力学习各种计算机语言，因此语言理解技术难度更大。

自然语言处理最大的挑战是处理歧义问题，多元化的人际交互需求倒逼语言处理技术不断升级。人类语言具有很大的复杂性，很多用词和表达方式与特定语境相关。即使是同一个句子，在不同语境下的意思也不相同。这需要计算机进行词法分析、语法分析、句法分析、语义理解和机器学习，是一项复杂的信息系统工程。

自然语言处理是包括人际交互在内的许多 AI 应用基础。目前的具体应用形态包括机器翻译、信息提取、报告生成、自然语言问答、聊天机器人等。如微软 Skype 提供语音-语言的实时翻译；Narrative Science 的 Quill 软件是一款机器人记者，可以从大数据中解读并撰写文章，实现拟人化的写作，已广泛应用于金融和新闻业；Apple、Facebook、微软纷纷发布聊天机器人；国内的小 i 机器人应用于各类企业的智能客服（VCA）。

（四）机器人=机械+人工智能

按照使用环境不同可分为两大类：工业机器人和服务机器人。工业机器人根据用途不同，大致可以分为焊接机器人、搬运机器人、喷漆机器人、装配机器人、处理机器人等。服务机器人则是除工业机器人之外的、用于非制造业并服务于人类的各种先进机器人，主要包括个人/家用服务机器人和专业服务机器人。

机器人产业市场潜力巨大。目前机器人在全球制造业密度只有 0.62%，我国仅有 0.3%，与世界平均水平相比还有一定差距，未来拥有巨大的市场发展空间。国内已经开始积极布局机器人产业，广东省发布《广东省机器人产业发展专项行动计划 2015-2017 年》，提出到 2017 年底，在 1950 家规模以上制造业企业开展工业机器人示范应用。富士康创始人也表示，在接下来几年里，其公司将会安装超过 100 万台的机器人。

七、人工智能的应用——为行业发展带来巨大变革

人工智能技术的不断发展为很多行业带来了再发展的难得契机，尤其以安防、汽车和医疗行业受到的影响最为深刻和直观。

（一）AI+安防：智能化推动安防行业的第三次升级

安防是人工智能在中国最容易变现的行业：交通拥堵及反恐等应用场景又急需最先进的人工智能技术。中国视频监控行业在过去十几年经历了两次重要的升级换代：

第一步，高清化：在这次升级的主要变化是摄像机的清晰度从标清（30 万像素）升级到 100 万像素或以上。图像传输方法从原本通过同轴电缆传输的模拟信号过渡到通过局域 IP 网或同轴电缆传输的数字信号。后端设备也从 DVR(Digital Video Recorder)过渡到 NVR (Network Video Recorder)。

第二步，网络化：在这次升级中的主要变化是，视频被直接传回数据中心内的集中存储(IP-SAN)。主要的优势是方便集中管理以及可监控的区域大大增加。随着2016年以来人工智能技术在视频分析领域的突破，视频监控行业正处在第三次重要的升级周期的开始阶段。

第三步，智能化：这次升级主要包括：(a)前端摄像机的智能化升级以支持结构化数据提取，(b)后端设备强化计算分析功能，以支持复杂的视频分析，(c)对应特定行业应用的人工智能分析软件快速增长。

1. 安防摄像机的智能化升级

传统的网络摄像机直接把高清视频回传给数据中心里的NVR，由于回传视频数据量巨大，很难对所有图像进行实时分析。大部分时候是对保存的图像进行事后分析。通过在网络摄像头上添加人工智能芯片(例如，NVidia的Jetson TX2、Movidius的Myriad2Vision等芯片)，前端摄像头可以实时对视频数据进行结构化处理。例如，设在交通路口的摄像头可以提取车牌车型等汽车信息、乘客数量、是否带安全带等乘客信息回传给数据中心。方便进行实时分析，优化系统反应能力。

根据统计，全球安防摄像头市场2016年约95亿美元，预计到2020年将达到128.4亿美元，CAGR为8.1%。其中，网络摄像机占比将从2016年的82%上升到2020年的90%。同时，相当一部分网络摄像机将迎来智能化升级。目前，中国占据全球44%的需求，海康威视已成为安防摄像头的全球龙头。

2. 后端设备强化计算分析功能

后端设备强化计算分析功能，以支持复杂的视频分析：传统的NVR(Network video recorder)的主要功能是压缩存储视频信息。通过添加GPU等人工智能加速芯片和应用处理软件，智能NVR能够实现图像识别，特征提取，人体识别、人员检索等功能。根据IHS的统计，后端录像存储设备的市场规模2016年约38亿美元。中国依旧占据了全球42%的市场需求，海康威视成为全球龙头。

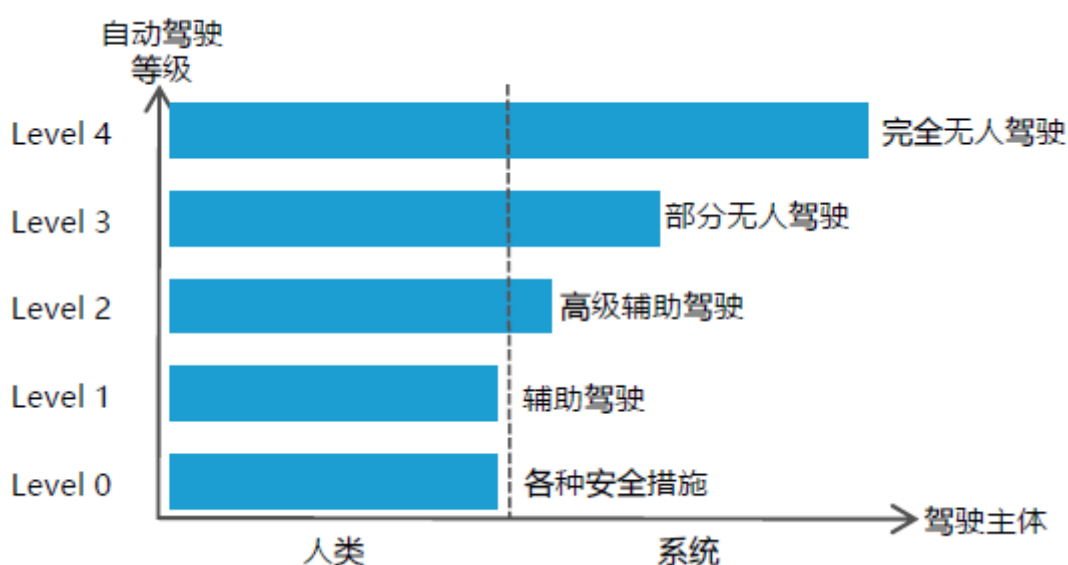
目前人工智能相关产品主要目标市场是政府市场。其驱动力来自于各地政府强化各类平安城市相关工程。经过十几年的平安城市建设，中国的主要城市已经积累了强大的视频数据采集能力。如何利用采集到的数据解决交通拥堵及治安问题一直是各地政府有待解决的问题。基于人工智能的新一代视频分析技术为城市

管理开辟了一条新的路径。并将驱动当前 67 亿美元的中国安防设备市场继续以年增速 15%以上增长。其中，交通管理和公共安全等政府相关项目（约占整体市场规模 30%）是最主要的应用领域。

（二）AI+汽车：有限无人驾驶体验提升，完全无人驾驶成为可能

AI 引领下，智能驾驶对交通运输和汽车制造业带来革新颠覆。智能驾驶作为人工智能落地最快、产业空间最大的应用方向之一，行业当下处于优势资源聚拢整合的阶段，整车厂商以及新兴科技巨头纷纷加大布局。以谷歌、百度以及初创科技公司为主的“越级式”阵营以及以传统汽车厂商为主的“递进式”阵营都怀抱终极愿景：“在任何区域达到完全无人驾驶”。

人工智能与汽车产业的融合，经历 ADAS 到 AV 的阶段。无人驾驶技术分五个阶段，依次从完全不具备自动驾驶功能的汽车，发展到完全自动驾驶汽车。L1 阶段和 L2 阶段为有限无人驾驶（ADAS），而 L3 阶段和 L4 阶段为完全无人驾驶（AV），当前绝大多数无人驾驶汽车尚处于 ADAS 阶段。这一过程的演进，是一个人类逐步交出操控权、提高安全系数的过程。



如上图所示，无人驾驶演进的阶段，是车辆操控权由人逐步交给计算机系统的一个过程。无人驾驶涉及的技术可以分为感知和决策两个层面，一方面通过传感器数据获取局部数据（车辆自身及四周环境的数据），另一方面结合高精度地图和天气数据做到构建全局数据。数据综合起来将与决策层做协调应用，辅助系统做定位和导航，再结合算法模型做路径规划，控制车辆的转向和速度，实现驾驶自动化。

除了传统汽车厂商，科技公司也把目光聚焦在无人驾驶领域，通常从汽车智能化的核心 软件技术入手，切入无人驾驶领域。百度和谷歌在高精度地图方面有显著优势，Uber 在 无人货运方面已有布局，苹果开发了智能防撞系统。

自动驾驶领域的合作趋势日趋明显。1) 对传统汽车厂商来说，与互联网公司、有科技含量的零部件公司、以及汽车共享服务商开展合作，是避免被淘汰的有效路径；2) 对科技企业来说，自主造车并非最明智的选择，毕竟未来汽车不仅仅等于“互联网+轮子”；3) 对零部件供应商来说，只有依托于汽车制造和科技企业，才能推动汽车互联、加速无人驾驶或自动驾驶的普及，从而创造价值。

（三）AI+医疗：从电子病历到辅助诊疗手术机器人，应用前景广阔

语音交互、计算机视觉、等技术的蓬勃发展，助推医疗领域的快速突破。通过人工智能的手段，医生诊断、患者自诊的效率大幅提升。具体的场景包括：语音录入病历、医疗影像智能识别、辅助诊疗/癌症诊断、医疗机器人等等。

1. 语音录入病历：国际巨头已现，中国市场发展空间广阔

传统电子病历录入耗时过长，成为医生的职业困扰。电子病历（EMR）的设计初衷是实现高效诊疗和数据共享，但实际操作中医生却需要耗费大量时间录入病历，诊疗效率并未得到有效提升。据调查，50%以上的住院医生每天用于写病历的平均时间达 4 小时以上，其中一部份甚至超过 7 小时；据美国医学会（AMA）统计，医生职业生涯约 15%~20%的时间用在病历书写等文档工作上，而接受训练的医生（如住院医师）所耗费的时间更是高达 30%。电子病历录入的简易性和高效性成为临床医生的迫切需求。

语音录入成为主要解决方案，兼容性和准确率是应用的关键。语音电子病历允许医生通过口述的方式进行病历的录入，成为代替传统电子病历的主要解决方案。语音输入系统需要与现有的医疗信息化系统相结合；此外，准确率是语音录入的核心评价指标之一，尤其医疗领域的专有名词数量众多，必须进行有针对性的优化，这也为准确识别和录入增加了难度。

（1）Nuance：基于语音技术精耕医疗领域，Dragon Medical 成为语音病历龙头

Nuance 是语音及语言解决方案行业的龙头公司，目前世界最先进的语音识别软件即出自 Nuance。公司主要经营医疗、企业、图形、移动 4 个业务板块，2014~2016 财年医疗板块的收入稳定在 10 亿美元左右，占总收入的比例约 50%。

据统计，Dragon Medical 平台的语音病历录入准确率已经高达 99%，能够帮助临床医生提升文档事务的效率达 45%。基于 Nuance 的临床语音识别技术，医生平均每年能够记录一亿病人的数据。目前，Nuance 的医疗语音解决方案在美国医疗机构中的覆盖率高达 72%，其客户分布在全球 30 余个国家和地区，已经有 50 万名临床医师和 1 万台医疗设备采用其医疗语音解决方案。

（2）科大讯飞：国内语音巨头布局医疗智能，语音病历产品初步上线

“基于语音的门诊病历采集系统”是讯飞医疗智能语音的主要产品之一。该系统由三部分组成：夹在医生领口的医学麦克风，装在医生口袋的发射器及插在工作电脑上的接收器。诊断过程中，医生以口述的方式说出患者的病历，讯飞通过自然语言处理技术将其转换成结构化的信息，形成包含患者检查史、病史、检查结果、身体指标等信息的结构化电子病历，并在工作电脑上生成记录，识别准确率高达 97%。

2016 年启动试点并初步应用，未来发展空间广阔。截至 2016 年 7 月底，我国共有医疗卫生机构 99.0 万个，其中包括 2.8 万家医院；按单一项目百万元的合同金额计算，潜在市场规模有望超过 500 亿元。2016 年 9 月，科大讯飞与北京大学口腔医院口腔数字化医疗技术和材料国家工程实验室共建的“基于语音的门诊病历采集系统”投入试点；截至 2016 年底，讯飞智能语音系统已经在 301 医院、瑞金医院、北京大学口腔医院、安徽省立医院等 20 多家医院应用；2017 年 4 月，讯飞与北京协和医学院签署战略合作框架协议，推进语音电子病历是该项合作协议的内容之一。相比 Dragon Medical 平台近 10 年的应用和 50 万名医师的覆盖，科大讯飞的语音病历系统仍处于商用的初级阶段，产品的推广和技术升级仍有待时日，但由于该产品并未进入中国市场，同为语音龙头的科大讯飞将具备更广阔的发展空间。

2. 医疗影像智能识别：上市公司和创业公司纷纷布局，整体处于商业化初期阶段

医疗影像数据丰富，人工识别存在不足。据统计，医疗行业 90% 的数据源于医疗影像，而现阶段的医疗影像诊断分析主要由医师人工完成，存在着一定的不足。（1）医师数量 存在缺口。医疗影像数据保持着年均 30.0% 的高速增长，而中国放射科医师数量年均增速仅 4.1%，基层医生更是严重缺乏，人手不足往往导致影像科医生工作时间过长；（2）医师的疲劳或经验不足可能引起错误的诊

断。据统计，我国普通疾病的误诊率高达 27%左右，重大疾病的误诊率更是高达 40%左右；（3）对影像的识别存在较强的主观性，医师之间的技术和经验有所差异，对同一资料的解读也可能存在分歧；（4）人工识别仅基于个人小样本的经验，很难利用全行业的大规模数据获取进一步的有效信息。

智能识别基于图形学和深度学习技术，有望提升信息复用率、识别效率和准确率。在医疗影像的智能识别中，首先借助计算机图形技术对医疗影像进行预处理、图像分割、特征提取、匹配判断等处理，而后结合后台数据库中的医疗影像数据，基于深度学习方式构建模型，针对医疗影像进行识别并给出判断。相比传统人工读片，智能识别充分提高了信息的重复利用率，能够以更客观、低成本、高效率的方式进行影像识别。

智能识别解决行业痛点，实现医院、患者、医生三方面共赢。医生方面，智能识别减少了读片时间，辅助医生降低误诊概率，提高诊疗水平；患者方面，智能识别能够有效减少患者的诊疗时间，方便患者进行远程诊疗，享受一线城市和大型三甲医院的高水平医疗；医院方面，对大规模的数据加以充分利用，建立整体的数字化平台，提高医院的核心业务能力，同时也有望推进医院之间的数据交换和共享。

医疗影像识别的主要难点在于数据获取、数据标注和跨学科人才积累。（1）数据获取：数据是深度学习算法所需的核心资源，仅掌握算法而缺乏数据无法获得较好的训练效果。现阶段，我国的医疗影像仍处于从传统胶片向电子数据过渡的阶段，大量影像资料尚未数字化，且医院之间的数据共享和互通程度较低，获取大规模的数据对业内公司是一个考验；（2）数据标注：在获取数据的基础上，深度学习结合先验知识对模型进行训练，训练集需要事先标注。由于大多数标注依赖人工识别，因此数据标注将耗费较大量人力和时间，在医疗影像领域获取具有高可靠性的标注数据也成为挑战之一；（3）“AI+医疗”跨学科人才积累：在较为专业的诊疗领域，应用及平台开发者不仅要研究人工智能算法，更要对医疗影像识别建立深入了解，人工智能+医疗的复合背景人才构成核心竞争力之一。

3. 辅助诊疗：以 Watson 为成熟案例，规模化应用值得期待

辅助诊疗作为认知智能的深层次应用，是“AI+医疗”的核心场景之一。通过理解和学习医师的专业知识，辅助诊疗技术平台模仿专业医师的判断过程，针

对疾病做出诊断，并给出诊断的置信范围。目前，辅助诊疗行业较为成熟的案例是 IBM 的 Watson 系统，国内人工智能公司科大讯飞也在积极布局。

IBM Watson: 辅助诊疗行业的最成熟产品，专注于癌症诊断领域。Watson 与纪念斯隆-凯特琳癌症中心共同训练其肿瘤解决方案（Watson for Oncology），基于 MSKCC 的数千份病人病历、近 500 份医学期刊、1,500 万页的医学文献等资料，Watson 被训练成了合格的肿瘤专家。2015 年 7 月 Watson 正式商用，它能够在诊断中能够分别给出推荐的、可考虑的和 not 推荐的三类治疗方案，并基于每种方案提供其逻辑、临床证据和用药信息。2016 年，美国北卡罗林那大学医学院的学者让 Watson 诊断了超过 1,000 例癌症病例，在超过 99% 的病例中 Watson 与肿瘤专家的治疗方案相同。IBM 全球副总裁陈黎明透露，甚至有一位东京女性已被医院认为治愈无望，但 Watson 在 10 分钟内阅读了 4,000 米厚的医学资料，诊断其患有一种罕见白血病，并给出了成功的治疗方案。

科大讯飞：差异化布局全科辅助诊疗，产品成熟仍需时间。以全科为定位，辅助基层医生诊疗。与 IBM Watson 不同，科大讯飞的辅助诊疗系统以基层医生为目标客户，希望在各科室的疾病诊断中都能够提供辅助建议，旨在缓解我国医生数量不足、基层医生水平有限等问题。科大讯飞已经开始将人卫出版社部分科室的电子书交给机器学习，也在系统中辅助增加了典型的病历分析。但辅助诊断系统的商业应用仍然还有很多工作要做，包括政策、产品流程及设计等。

4. 医疗机器人：手术机器人达芬奇普及度较高，导诊机器人集中面市

按照应用场景，医疗机器人可分为手术机器人、诊疗机器人、护理机器人、康复机器人、导诊机器人、医用教学机器人等。手术机器人是整个医疗机器人市场中占比最大的产品。

（1）手术机器人

达芬奇机器人垄断微创手术，国产骨科机器人“天玑”世界领先。达芬奇机器人垄断微创手术，在全球广泛应用。达芬奇手术机器人 2000 年正式面市，已推出了 4 代产品。机器人由外科医生控制台、床旁机械臂系统、成像系统三部分组成，通过微创方法实施复杂的外科手术，主要应用在泌尿外科、普外科、胸外科、妇产科等科室。达芬奇机器人系统和耗材成本高昂，对其在中国的推广形成阻碍。从系统售价来看，达芬奇机器人手术系统 Xi（第四代产品）在中国的售价约 2,000~2,500 万元，远高于美国的售价 150 万美元。此外，其组件机械臂是一种耗材，每条机械臂使用次数不得超过 10 次，每场手术需同时使用 3~4 条机

械臂，其价格约为 10 万元/条。以此测算，平均每台手术的价格比传统手术高 3~4 万元。

我国自主知识产权的“天玑”骨科机器人已经达到世界领先水平。“天玑”机器人由北京积水潭医院、北京航空航天大学 and 北京天智航医疗科技三方联合研发，于 2016 年正式获得 CFDA（国家食品药品监督管理总局）颁发的医疗器械注册证。天玑机器人采用计算机辅助导航技术解决视野问题，用机械臂定位解决高精度操作问题，能够达到 0.8 毫米的操作精度，是目前国际上唯一可以开展全节段脊柱手术的机器人系统。临床试验中，天玑机器人在脊柱手术中导针植入精度可提高 1 倍，植入时间减少 25%；在创伤手术中螺钉位点优良率达 100%，X 线透视时间减少 83%，导针植入一次性成功率 90%以上，技术数据达到世界领先水平。

（2）其他类型医疗机器人

2017 年集中面市，有望缓解导诊高峰的人手不足 2017 年以来，导诊机器人集中面市。随着 2017 年 3 月科大讯飞的“晓曼机器人”在合肥第一人民医院成为首个投入使用的导诊机器人，小胖、阿鹏、糖宝等机器人相继开始应用。科大讯飞的晓曼机器人是可以应用于各个领域的服务机器人。它集成了最新的科大讯飞 AIUI 技术，可以结合人脸、声纹、远场识别等技术，完成用户身份识别、数据分析、业务办理、语音问答等工作，在金融、政府、医疗等各领域都具有大量的应用场景。2017 年以来，合肥市第一人民医院和北京 301 医院先后引入晓曼机器人进行智能导诊，使用者可以通过语音问答、触屏输入等互动方式，进行医院位点咨询、219 个常见病和症状咨询、51 个常见知识问询等。此外，晓曼机器人有能力通过自适应的学习提升在方言上的理解能力，不断提升方言语音问答的效果。导诊机器人的应用有望缓解医院问询台高峰期人手不足、等待时间长、效率低下的现状，提升医患双方的工作和就医体验。

医疗行业空间巨大，但技术还有待成熟。电子病历的建立，不仅仅用到了语音识别技术，也整合了医疗大数据；影像诊断则用到了图像识别技术，现在都已经形成成熟的商业模式。辅助治疗和手术机器人由于技术的尚不完善，还在小范围推广。由于语音识别技术和机器视觉技术的成熟，大量应用这两个技术的行业将迎来 AI 变现的更多机会。