

人工智能指数 2018 年度报告

【译者按】 2018年12月，斯坦福大学、麻省理工学院与 OpenAI 联合发布了《人工智能指数 2018 年度报告》。本期指数报告是继 2017 年始发之后的第二期，包括活动量和技术性能两个核心指标以及若干其它指标，详细分析了人工智能在学术研究、行业、开源软件、公共事务等方面的发展，阐述了各项技术性能的进展情况，并介绍了主要国家和地区人工智能近期计划。报告认为，近年来中国在人工智能领域发展迅速，但美国在人工智能论文引用、专利、投资等方面仍处于世界领先地位。报告还指出，自然语言处理领域的性能得到显著改善。赛迪智库规划研究所对该报告进行了编译，期望对我国有关部门有所帮助。

【关键词】 人工智能指数 年度报告 2018

由斯坦福大学、麻省理工学院与 OpenAI¹联合发布的人工智能指数 2018 年度报告相比 2017 年报告有以下两方面的改进：一是基于对人工智能领域相关活动的持续跟踪，更新了最新指标；二是站在全球化视角进行了解析。2017 年的指数报告涉及较多北美区域的活动，主要是因为当时全球只有为数不多的几个经济体就人工智能项目建立了合作关系。然而，当前人工智能已经在全球多个国家快速发展，主要体现在：一是人工智能论文呈现多极化趋势。如 2017 年 Scopus 数据库中有 83% 的人工智能论文来自美国以外的地区，其中，28% 来自欧洲地区。二是人工智能教育呈现泛化趋势。如全球注册人工智能（AI）和机器学习（ML）课程的人数在不断增加，尤其是在中国清华大学，其相关注册人数比 2010 年注册人数增加了 16 倍之多。三是多个国家或地区的人工智能相关专利快速增长。除美国、中国和欧洲外，其他经济体也取得了进展。2014 年，韩国和日本分别是第二和第三大人工智能专利申报国，仅次于美国。

一、指标说明

（一）核心指标

2018 年人工智能指数有两个核心指标：活动量指标和技术性能

¹ OpenAI 是由诸多硅谷大亨联合建立的人工智能非营利研究组织。

指标。活动量指标主要用于衡量学者、企业、企业家以及公众在人工智能领域的参与度，具体数据包括学习人工智能的本科生人数、申请人工智能工作的女性人数占比以及创办人工智能企业所需风投资金的增长率等。技术性能指标主要用于衡量人工智能细分领域的技术性能变化情况。新版的人工智能指数新增了国家或地区型指标，如国家层面的机器人研发管理机构、人工智能会议参会者等指标。总体而言，上期报告所体现的主要趋势仍在延续，即人工智能活动正在向世界各地普及，且技术性能正在全面提升。

（二）其他指标

本期报告在“衍生指标”章节重点分析了发展趋势之间的关系，探索出一项新的指标即人工智能活力指数。该指数通过综合学术界和行业界的趋势，从而量化人工智能领域的活力情况。本期报告还引入一个新的定性指标，即政府近期计划。该指标可用于体现政府近期对人工智能的投资计划。本期报告重点对美国、中国和欧洲的近期计划进行了研究，未来将通过与其他机构合作，对更多的国家和地区人工智能政府计划进行分析。

二、核心指标：活动量

（一）学术研究

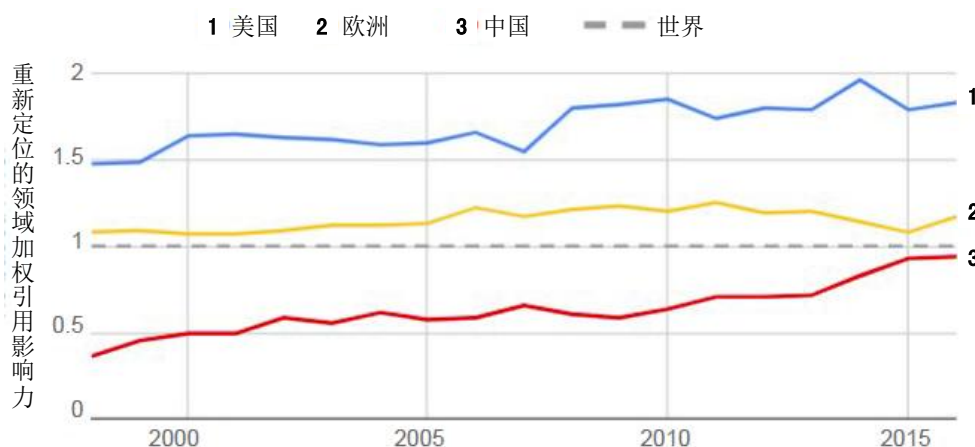
1、发表论文

从论文发表数量看，以斯高帕斯数据库中论文为例，人工智能论文数比 1996 年增加了 8 倍，计算机科学论文数比 1996 年增加了 6 倍。从发表区域看，2017 年，斯高帕斯数据库中有 28% 的人工智能论文来自欧洲地区，其次是中国（25%）和美国（17%）。从发表主题看，2017 年，机器学习与概率推理类的论文数占比为 56%，而 2010 年仅为 28%。2014-2017 年间，大多数类别论文的发表速度高于 2010-2014 年。尤其值得注意的是，2010-2014 年，神经网络论文的年复合增长率（CAGR）仅为 3%，而 2014-2017 年的年复合增长率为 37%。从各地区相对活动指数（RAI）²看，中国的人工智能论文主要集中在工程技术和农业科学领域，而美国和欧洲的人工智能论文则主要集中在人文以及医疗与卫生科学领域。2017 年数据显示，与 2000 年相比，上述三个地区对人工智能研究的重视程度均有所提高，且中国正转向农业领域的人工智能研究。

从论文发表主体看，2017 年中国政府机构人工智能论文发表量是企业发表量的 4 倍。自 2007 年以来，中国政府机构发表的人工智能论文的数量增加了 400%，而企业论文发表量仅增加了 73%。在美国，企业发表的人工智能论文占比相对较大。2017 年，

² 相对活动指数是指一国人工智能论文发表量在本国论文发表量中的占比与全球人工智能论文发表量在全球论文发表量中的占比的比值。若该值为 1.0，则表示该国的人工智能研究活动与全球人工智能研究活动完全一致；若该值大于 1.0，则表示该国更加注重人工智能方面的研究；若该值小于 1.0，则表示该国对人工智能研究的重视程度低于全球水平。

美国企业人工智能论文发表量占本国所有人工智能论文发表量的比例比中国高出 6.6 倍，比欧洲高出 4.1 倍。此外，美国企业人工智能论文发表量增幅最大，其 2017 年企业人工智能论文发表量是 2009 年的 1.7 倍。从论文引用影响力³看（见图 1），虽然欧洲每年发布的人工智能论文数量最多，但其重新设定的领域加权引用影响力仍保持相对平稳，与世界平均水平相当。相比之下，中国重新设定的领域加权引用影响力大幅提升。2016 年，中国人工智能作者的被引用次数比 2000 年高出 44%。然而，在总引用次数方面，美国仍优于其他地区。美国作者的被引用次数比全球平均水平高出 83%。



来源：爱思唯尔（Elsevier）信息分析公司

图 1：各地区人工智能作者的领域加权引用影响力（1998-2016 年）

³ 引用影响力 (FWCI) 是指该地区人工智能作者平均被引用次数与全球所有人工智能作者平均被引用次数的比值。该定义重新设定了领域加权引用影响力的基础，即引用影响力将以相对于世界平均水平的比值进行显示。若领域加权引用影响力为 1，则表明该地区出版物的引用次数与世界平均水平相当；若领域加权引用影响力为 0.85，则表明该地区出版物的引用次数比世界平均水平低 15%。

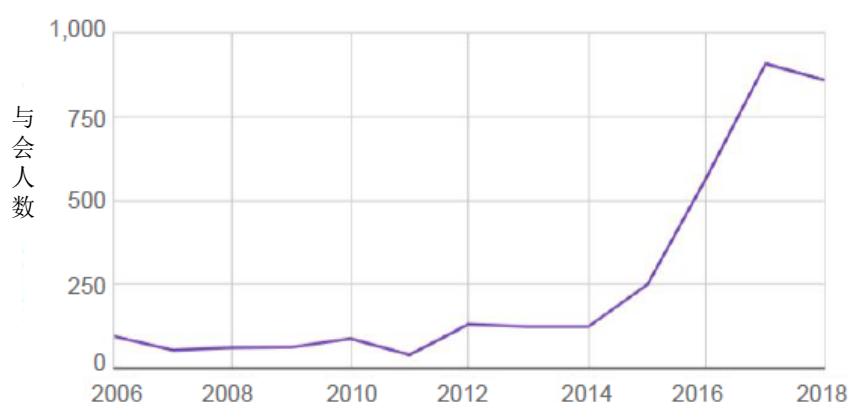
2、课程注册

以美国人工智能课程为例，美国几所领先的计算机科学大学的人工智能和机器学习课程的注册人数正在不断增长。2017年，入门级人工智能课程的注册人数比2012年增加了3.4倍，而入门级机器学习课程的注册人数则比2012年增加了5倍。2017年，加州大学伯克利分校的入门级机器学习课程的注册人数是2012年的6.8倍，这一增长率高于其他所有被调研的学校。以国际其他高校课程来说，2017年，中国的清华大学人工智能和机器学习课程的注册人数比2010年增加了16倍。多伦多大学、维也纳技术大学、不列颠哥伦比亚大学等人工智能和机器学习课程的注册人数都逐年增长。总的来看，在所有研究的学校中，人工智能课程注册人数增长率在一定程度上取决于学校，而地理位置并不会造成太大影响。

3、参与度

参与度指标主要包括参加人工智能会议和注册各种人工智能相关的机构。数据显示，2017年参与大型（人数超过2000人）人工智能会议的人数增长迅速，其中，神经信息处理系统会议和国际机器学习会议的与会人数增长速度最快，分别是2012年与会人数的4.8倍和6.8倍。参与小型（人数不足2000人）人工智能会议的人数也有显著增长，特别是国际学习表征会议(ICLR)，

2018 年的与会人数比 2012 年增加了 20 倍。注册女性机器学习研讨会和 AI4ALL⁴等机构的人数均有所增加，女性机器学习研讨会的与会人数比 2014 年增加了 6 倍，AI4ALL 的毕业人数比 2015 年增加了 9 倍（见图 2）。这些增长表明，当今社会一直在为提高女性和非代表性群体在人工智能领域的参与度而努力。



数据来源：女性机器学习研讨会

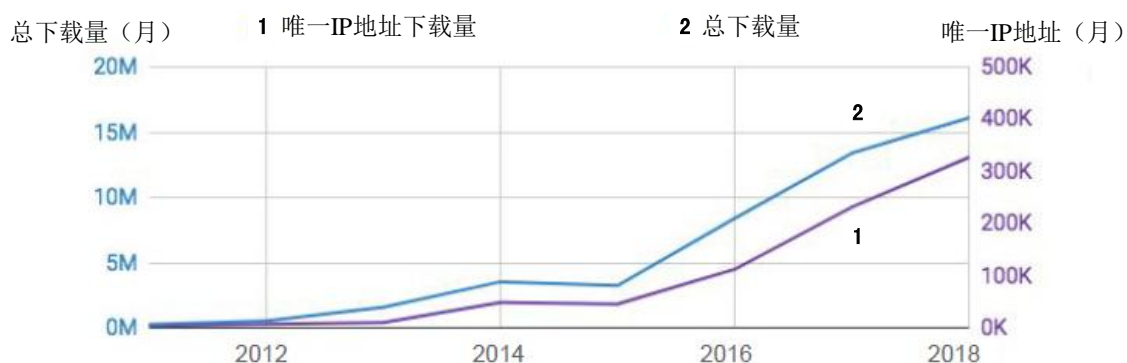
图 2：女性机器学习研讨会登记人数（2006–2018 年）

4、机器人软件下载

机器人操作系统是一种广泛使用的机器人开源软件栈，许多商业制造商和学术研究人员都使用机器人操作系统。图 3 显示了 ROS.org 网站上机器人操作系统（ROS）二进制包下载量随时间的变化情况，左轴代表月平均下载量，右轴代表仅来自唯一 IP 地址的月平均下载量。自 2014 年以来，总下载量和唯一 IP 地址下

⁴ 一个致力于提高人工智能教育、研究、发展和政策方面的多元化与包容性的非营利组织，其通过人工智能夏令营活动支持下一代多元化的 AI 人才，并对年轻女性、代表性不足的群体和社会经济地位较低的家庭做出特殊考虑。

载量分别增长了 352% 和 567%。这就表明，当今社会对机器人技术以及机器人系统的应用越来越感兴趣。目前，唯一 IP 地址下载量的增长速度高于总下载量，由此可以推断，机器人操作系统用户也有所增加，而不仅仅是使用频率的增加。从地区上来看，美国和欧洲的 ROS.org 页面浏览量最多，中国紧随其后，且中国是增长率最高的地区。



来源：机器人操作系统度量报告

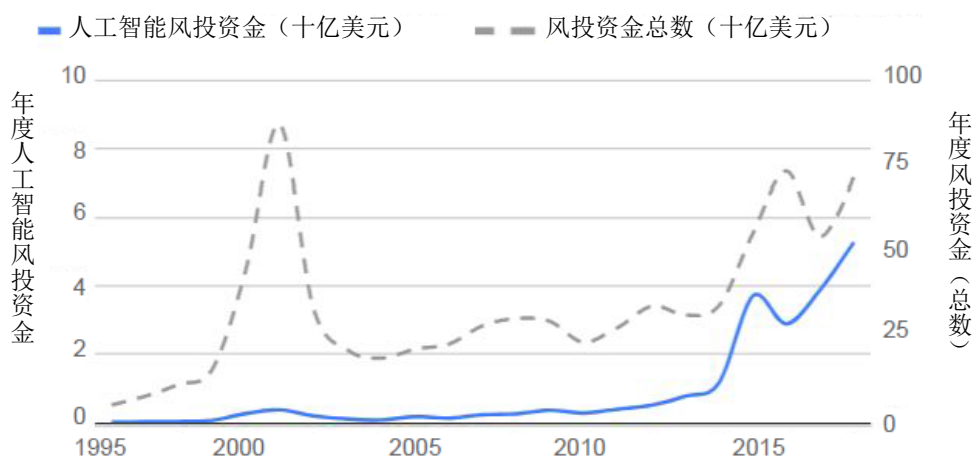
图 3：机器人操作系统软件包下载量（2011-2018 年）

（二）行业状况

1、创业/投资

从人工智能初创企业来看，到 2018 年 1 月，美国由风险资本支持且正常经营的私营初创企业超过了 10000 家，比 2015 年 1 月增加了 2.1 倍，呈现指数级增长态势。从风险投资资金来看，2017 年，美国风投资金总额比 2013 年增加了 2.08 倍，其中人工

智能领域的风投资金比 2013 年增加了 4.5 倍。图 4 显示了风投公司在所有融资阶段向正常经营的美国初创企业投入的年度融资金额，1997-2000 年，网络泡沫导致风投资金激增，而 2014 和 2015 年风投资金的小规模增长则反映在此期间相对较大的经济增长。



来源：山德希尔咨询公司

图 4：美国人工智能初创企业年度风投资金（1995-2017 年）

2、就业

就业方面则呈现两大突出问题，一是技能岗位空缺，二是申请人性别差异较大。从岗位来看，机器学习技能的需求量最大，深度学习技能需求量的增长速度最快。2017 年，需要深度学习技能的岗位空缺数量比 2015 年增加了 35 倍。从申请人性别差异来看，平均而言，美国人工智能岗位申请人中男性占比 71%。由于机器学习岗位的申请人数最多，因此，该类岗位申请人中的性别

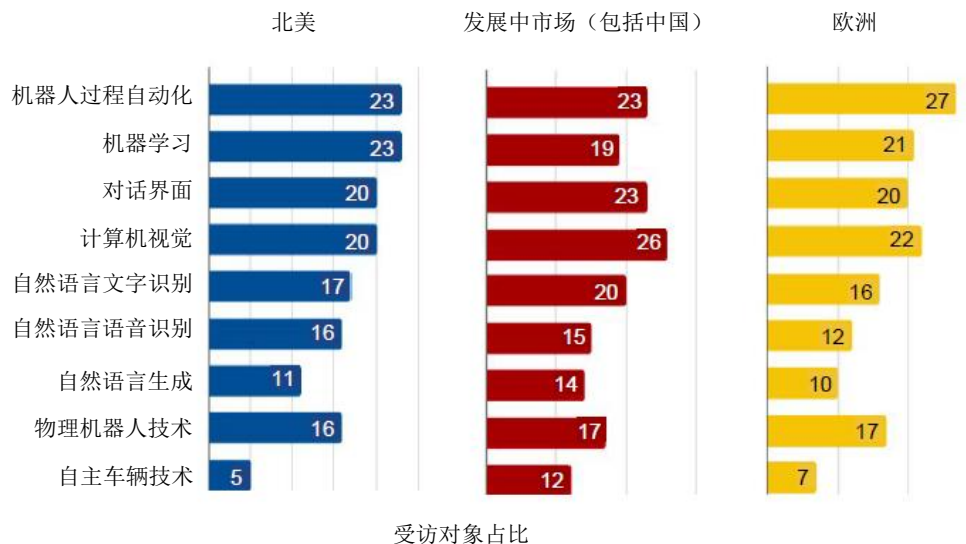
差异是导致这一平均水平的主要原因。此外，相对于其他类别的人工智能技术，深度学习和机器人技术岗位的申请人也存在较大的性别差异。

3、专利

从各发明地区人工智能专利的数量和增长情况来看，2014年，大约30%的人工智能专利集中在美国，其次是韩国和日本，占比均为16%。在发明量最多的地区中，韩国和中国台湾增长最快，其2014年人工智能专利数量几乎是2004年的5倍。

4、人工智能采用

根据麦肯锡咨询公司对2135名不同机构受访对象展开的调查，图5显示了至少已在其一个功能或业务部门引入人工智能能力的受访对象占比情况。分地区来看，虽然有些地区某种能力的采用程度远高于其他地区，但在人工智能能力方面，各地区的采用程度大致相等。分领域来看，各行业纷纷引入人工智能能力以获得最大行业价值。例如，金融服务业在风险功能中大量引入人工智能，汽车行业在制造功能中大量引入人工智能，而零售业则在营销/销售功能中大量引入人工智能。这就意味着，特定应用（如制造）的人工智能的发展速度可能与特别强调专业化的行业有关。



来源：麦肯锡咨询公司

注：柱形图的大小是以各地区的能力为基础；北美：N=479；发展中市场（包括中国）：N=189（中国：N=35）；欧洲：N=803

图 5：至少已在一项企业功能中引入人工智能能力的受访对象占比（2018 年）

5、财报电话会议中的提及次数

各行业的企业财报电话会议中人工智能和机器学习的提及次数整体呈现增长态势，其中，IT 企业提及的次数最多。2015 年，IT 行业提及人工智能和机器学习的次数开始大幅增加，而对其他大多数行业而言，2016 年才开始出现此类增长。IT、非必需消费品、金融和医疗行业在财报电话会议中提及人工智能的次数最多。

6、机器人机构

从各地区工业机器人机构的年度数量来看，自 2012 年以来，

中国机器人机构年度数量已增加了 500%，其他地区也有显著增长，如韩国和欧洲，分别增加了 105% 和 122%。目前，在数量较少的地区中，中国台湾的年度机构数量最大，且在 2012-2017 年间增长最快，机器人机构总数已超 10000 家。

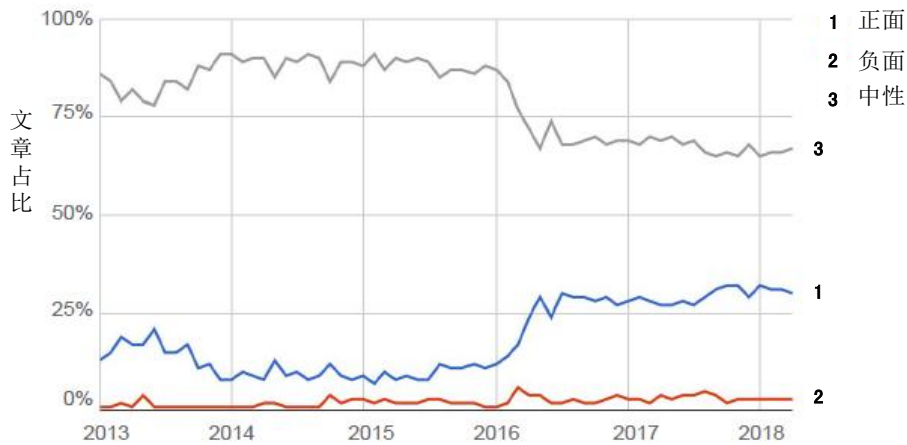
（三）开源软件

开源软件有两个最新趋势，一是大型企业支持的框架，如谷歌的 Tensorflow、脸书的 Pytorch、亚马逊的 mxnet 等，正在日益普及；二是相对于其他编程语言，TensorFlow 最为普及。

（四）公共事务

1、媒体态度

从媒体对人工智能的评判态度来看，中性评判人工智能类文章越来越少，正面评判人工智能类文章越来越多。自 2016 年初期以来，2016 年 1 月至 7 月，正面文章占比从 12% 增至 30%，此后，正面文章占比一直在 30% 左右（见图 6）。



来源：TrendKite 公司

图 6：人工智能类文章的态度（2013-2018 年）

2、政府提及率

自 2016 年以来，美国、加拿大和英国三国政府对人工智能和机器学习相关词汇的提及率不断攀升。在此之前，这三个国家鲜少提及机器学习。以美国为例，美国国会对人工智能和机器学习的提及率在 2018 年有了显著增长，美国数据中的每项计数代表至少提及过一次机器学习或人工智能的特定事件或对话。

三、核心指标：技术性能

（一）视觉

1、对象检测

ImageNet 挑战赛是全球计算机对象检测领域的重要竞赛，能够反映该领域的进展。从准确度方面看，根据 2012-2018 年

ImageNet2012 数据集的测试和验证结果，对象检测领域的性能不断提升，测试集和验证集准确度都已达到 95% 以上。从训练时间看，2017 年 6 月至 2018 年 11 月间，训练对象检测人工神经网络所需的时间已从 1 小时缩短至 4 分钟左右，训练速度提高了约 16 倍。其中，算法创新、基础设施投资是训练时间缩短的主要原因。

2、实例对象分割

由于对象检测与图像分类领域已经实现了较高性能，ImageNet 挑战赛已宣告结束。自 2017 年来，研究者开始将研究重点转向需要更复杂推理的视觉任务，如定位像素级精度的对象（实例对象分割）、将场景划分为具有像素级精度的区域（语义分割）等。COCO 挑战赛是全球实例对象分割的重要竞赛，其宗旨是构建一种可精确定位各对象并描绘其边界的算法。自 2015 年以来，COCO 挑战赛中达到的最高平均精度提高了 0.2 个百分点。

（二）语言

1、句法分析

句法分析往往是进行某些自然语言处理任务的第一个步骤。目前几乎普遍使用深度学习技术进行句法分析。2003 至 2018 年间，句法分析的性能提高了约 10%。

2、机器翻译

目前，机器翻译的性能有大幅提升。2018 年，BLEU 模型⁵在

⁵由 IBM 提出的评测机器翻译译文与参考译文之间相似度的算法。

英语译德语方面的得分比 2008 年提高 3.5 倍，而在德语译英语方面的得分比 2008 年提高 2.5 倍。

3、回答问题

AI2 推理挑战 (ARC)⁶。AI2 推理挑战数据集包含 7787 个小学水平 (美国 3-9 年级) 的多项选择题，这些问题可分为挑战题集 (2590 个问题) 和简易题集 (5197 个问题)。2018 年 4 月至 11 月期间，AI2 推理挑战赛中的简易题集性能得分由 63% 提升至 69%，挑战题集得分由 27% 提升至 42%。

通用语言理解评估 (GLUE)⁷。通用语言理解评估基准是一项新基准，旨在通过一系列任务测试自然语言理解 (NLU) 系统，并促进非特定任务系统的开发。该基准由九项子任务组成：两项单句任务 (衡量语言可接受性和情绪)，三项相似性与释义任务，以及四项自然语言推理任务。语料库大小从 1000 到 400000 不等。评估标准主要有准确度⁸/F1⁹和 Matthews 相关性系数¹⁰。尽管评估基准于 2018 年 5 月发布，但到 2018 年 10 月，性能已经由最初公布基线 (约 70%) 提高至约 80%。

⁶由非营利计算机实验室-艾伦人工智能研究所(AI2)设计的一项新测试——弧理论挑战(ARC)，用以测试人工智能，了解其对世界运行方式的理解。

⁷一种用于评估和分析多种已有自然语言理解任务的模型性能的工具，模型基于在所有任务的平均准确率进行评估。

⁸对于给定的测试数据集，分类器正确分类的样本数与总样本数之比。

⁹精确值与召回率的调和均值。

¹⁰综合考虑 TP、FP、TN、FN，对分类器性能进行综合评估，计算方法为 $(TP*TN - FP*FN) / \sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$ ，其中，TP 为真阳性、TN 为真阴性、FP 为伪阳性、FN 为伪阴性。

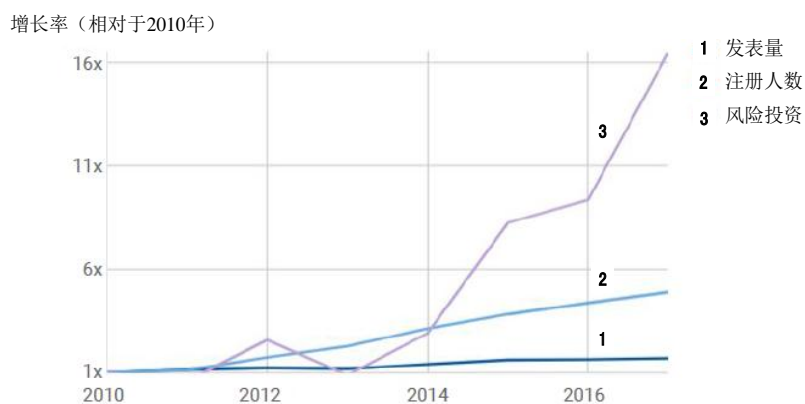
四、其他指标

（一）衍生指标

第一项衍生指标是学术界-行业界动态，该指标可体现所选学术指标以及行业界动态的增长情况。第二项衍生指标是人工智能活力指数，该指标是学术指标和行业指标的综合体现。

1、学术界-行业界动态

为探索学术界与行业界人工智能相关活动之间的关系，本报告首先从上述章节中选取了一些有代表性的衡量指标，其中包括斯高帕斯数据库中人工智能论文的发表量、美国几所大学入门级人工智能和机器学习课程的综合注册人数以及对人工智能初创企业的风险投资。这些指标体现的表征量无法直接进行对比。为分析各趋势之间的关系，本报告规范了从 2010 年开始使用的衡量指标，并显示了各指标的增长情况，而非绝对数量（见图 7）。



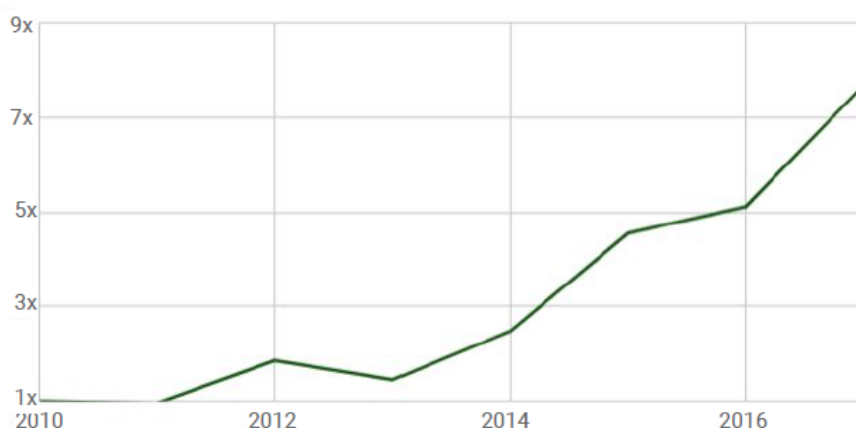
来源：山德希尔咨询公司、斯高帕斯数据库、大学提供的数据

图 7：学术界-行业界动态（2010-2017 年）

2、人工智能活力指数

人工智能活力指数是学术界-行业界三个衡量指标（论文发表量、课程注册人数和风险投资）的综合体现，本报告利用该指数衡量人工智能领域的活力。人工智能活力指数主要受风险投资的影响，因为相对于其他两项指标，风险投资增幅最大（见图 8）。

活力指数（相对于2010年）



来源：山德希尔咨询公司、斯高帕斯数据库、大学提供的数据

图 8：人工智能活力指数（2010-2017 年）

（二）政府近期计划

以下内容由欧洲委员会联合研究中心编写而成，其中也包括蒂姆·达顿（Tim Dutton）在 2018 年 6 月《国家人工智能战略概述》中所述内容（见表 1）。

表 1：主要国家和地区制定的人工智能相关战略

美国	(仅涵盖欧盟、英国和法国, 不包括其他欧洲国家的计划)	中国
<p>2016年10月：奥巴马政府提出了一项国家人工智能研发战略，目标是投资研究、开发人类-人工智能协作方法，应对人工智能带来的安全、道德、法律和社会影响，为人工智能培训创建公共数据集，并通过各种标准和基准评估人工智能技术。美国政府还发布了第一份关于人工智能的政策报告：《为人工智能未来做好准备》。</p> <p>2018年5月：特朗普政府举办的人工智能峰会宣布了其目标：（1）保持美国在人工智能领域的领导地位；（2）支持美国工人；（3）促进公共研发；（4）消除创新障碍。人工智能特别委员会专门负责向白宫提出有关人工智能研发重点的建议，并思考如何与学术界和业界建立伙伴关系。</p> <p>2018年9月：美国国防部高级研究计划局（DARPA）宣布了一项 20 多亿美元的投资计划，以攻克人工智能技术限制。“下一代人工智能”（AINext）项目启动。美国众议院监督与政府改革委员会信息技术小组委员会发表了一份有关人工智能及其政策影响的白皮书：《机器崛起：人工智能及其对美国政策日益凸显的影响》。</p> <p>私营企业在美国人工智能开发/投资中发挥着核心作用。2017年，亚马逊、Alphabet 等私营科技公司分别投入了 161 亿美元和 139 亿美元用于研发人工智能。相比之下，2019 年预算中，美国国家科学基金会总预算，加之美国国防部高级研究计划局和美国运输部（DOT）在自主和无人系统方面的投资，共计仅 53 亿美元。</p>	<p>欧洲</p> <p>2018年4月：欧盟成员国签署了《人工智能合作宣言》，同意一起合作。此外，欧盟委员会（EC）还发布了一份《人工智能通报》，其目标是：（1）促进欧盟的研究投资；（2）为社会经济变革做好准备；（3）建立道德和法律框架。该通报还投入了 15 亿欧元（17 亿美元），用于支持 2018-2020 年期间的人工智能研究。此外，还有来自“地平线 2020”计划的 26 亿欧元。欧盟委员会的目标是在未来 10 年内投入 200 亿欧元进行人工智能研究。</p> <p>2018年6月：欧盟委员会提出了一项“数字欧洲”（Digital Europe）计划，其 2021-2027 年间的预算为 92 亿欧元（104 亿美元）。该计划旨在推进人工智能技术的发展，并确保人工智能在整个经济和社会中的应用。欧盟委员会还提议开发所有人都可使用的“欧洲图书馆”通用算法。</p> <p>法国</p> <p>2018年：法国发布了三份最新报告：（1）《法国人工智能》计划，提出了研究、教育、创新战略；（2）科学与技术议会办公室编写的一份报告，重点关注社会和监管问题；（3）Mission Villani 报告，重点关注政策和人才流失问题，同时也指出了多元化人工智能研究的重要性。</p> <p>英国</p> <p>2016年：英国国会上议院人工智能特别委员会开始发布一系列人工智能政策报告。</p> <p>2018年4月：英国政府发布了《人工智能领域行动》（AI Sector Deal），该行动投资 9.5 亿英镑（12 亿美元）用于支持研究/教育，并加强英国的数据基础设施。</p>	<p>自 2014 年以来，中国政府推出了一系列重要的国家人工智能计划，其目标是到 2018 年在中国建立一个价值 147 亿美元的人工智能市场，并确保中国到 2030 年成为人工智能领域的全球领导者。</p> <p>2015年7月：中国“互联网+”计划的重点是智能制造。特定人工智能目标包括：增加公众对人工智能开发的支持；推动人工智能在智能家居、智能汽车和机器人等领域的普及；建立大型培训数据库，包括语音、图像、视频、地图和研究；开发关键人工智能技术并使其产业化，如计算机视觉、语音处理、人机交互等。</p> <p>2016-2020年：机器人产业发展计划为中国智能工业与服务机器人的发展指明了道路。</p> <p>2017年7月：“新一代人工智能发展计划”制定了到 2030 年人工智能发展战略。计划到 2020 年，中国的人工智能技术将达到全球领先水平；到 2025 年，中国将在人工智能理论方面取得重大突破，且人工智能将推动产业升级和经济体制改革；到 2030 年，中国将成为世界主要的人工智能创新中心。</p>

五、达到人类水平表现的大事件

2017年的《人工智能指数报告》包含“达到人类水平表现的大事件”章节，给出了人工智能达到或超过人类表现的各项事件及时间，包括游戏成就、医疗诊断以及其它复杂的人工任务。本期报告中，新增了四项重大事件（见表2）。

表 2：达到人类水平表现的大事件（1980–2018 年）

时间及事件	事件简介
1980 年 奥赛罗	20 世纪 80 年代，李开复和桑乔伊·马哈詹（Sanjoy Mahajan）开发了 BILL——一个用于玩奥赛罗棋盘游戏的贝叶斯学习系统。1989 年，该程序在美国电脑棋手大赛中赢得冠军，并以 56:8 的成绩击败了排名最高的美国选手布莱恩·罗斯。1997 年，Logistello 程序在与奥赛罗卫冕世界冠军的六场比赛中均取得了胜利。
1995 年 跳棋	1952 年，亚瑟·萨缪尔斯（Arthur Samuels）编写了一系列可玩西洋跳棋并通过自我对弈不断改进的程序。然而，直到 1995 年，Chinook 跳棋程序才赢得了世界冠军的称号。
1997 年 国际象棋	20 世纪 50 年代，部分计算机科学家预测，计算机将在 1967 年之前击败人类国际象棋冠军，但直到 1997 年，美国国际商用机器公司（IBM）的 DeepBlue 系统才击败国际象棋冠军加里·卡斯帕罗夫（Gary Kasparov）。目前，智能手机上的象棋程序可达到大师级水平。
2011 年 Jeopardy!	2011 年，美国国际商用机器公司的沃森（Watson）电脑系统在一个著名的智力竞赛电视节目（Jeopardy）上击败了前获胜者布拉德·拉特（Brad Rutter）和肯·詹宁斯（Ken Jennings），并赢得了 100 万美元的一等奖奖金。
2015 年 雅达利（Atari） 游戏	2015 年，谷歌 DeepMind 团队利用增强式学习系统学习如何玩 49 种雅达利游戏。该系统在大多数游戏（如 Breakout）中的表现可达到人类水平，但在某些游戏（如“蒙特祖马的复仇”（Montezuma's Revenge））中仍然表现欠佳。
2016 年 ImageNet 中的 对象检测	2016 年，ImageNet 的自动标注误差率从 2010 年的 28% 降至 3% 以下，而人力误差率为 5% 左右。
2016 年 围棋	2016 年 3 月，谷歌 DeepMind 团队开发的 AlphaGo 系统以 4: 1 的成绩击败了世界上最伟大的围棋选手之一李世石（Lee Sedol）。随后，DeepMind 发布了 AlphaGo Master 系统，该系统于 2017 年 3 月击败了排名最高的玩家柯杰（Ke Jie）。2017 年 10 月，英国《自然（Nature）》期刊上有一篇文章详细介绍了另一新版本系统——AlphaGo Zero，该系统以 100: 0 的完美成绩击败了最早的 AlphaGo 系统。

2017 年 皮肤癌分类	在 2017 年《自然》期刊的一篇文章中，爱斯特瓦（Esteva）等人描述了一个根据 2032 种不同疾病的 129450 张临床图像数据集进行训练的人工智能系统，并将其诊断性能与 21 名经过委员会认证的皮肤科医生进行对比，发现人工智能系统的皮肤癌分类能力与皮肤科医生水平相当。
2017 年 Switchboard 语 音识别	2017 年，微软和 IBM 在交换机（Switchboard）有限领域语音识别方面几乎实现了与人类水平相当的性能。
2017 年 扑克牌游戏	2017 年 1 月，卡耐基梅隆大学的 Libratus 程序在一场涵盖 12 万局双人、单人无限注德州扑克比赛的锦标赛中击败了四名顶级人类选手。2017 年 2 月，阿尔伯塔大学的 DeepStack 程序与 11 名专业选手分别进行了 3000 多局比赛。DeepStack 的获胜次数足以证明其技能明显优于专业人士。
2017 年 吃豆小精灵游戏	微软收购的深度学习团队 Maluuba 创建了一个人工智能系统，学习如何在雅达利 2600 游戏中取得最高分（999900 分）。
2018 年 汉英翻译	在将新闻报道从汉语译成英语方面，微软机器翻译系统可达到人类水平的翻译质量和准确度。该测试是基于 newstest2017 数据集进行的，该数据集是机器翻译竞赛中的常用数据集。
2018 年 “夺旗”游戏	DeepMind 计算机在改进版“雷神之锤 III 竞技场夺旗”（一种流行的 3D 多玩家第一人称视频游戏）中的表现达到了人类水平。游戏中，智能体表现出了类人行为，如导航、跟踪、防守等。无论是作为队友还是对手，受过训练的智能体的获胜率均超过了强大的人类选手，击败了多个现有的最新系统。
2018 年 魔兽争霸 (Dota) 2	在魔兽争霸 2 中，OpenAI 由五个神经网络组成的团队——OpenAI Five ¹¹ 击败了业余人类团队。OpenAI Five 每天进行自我对抗的游戏次数相当于 180 年的游戏量，以此方式进行训练（目前，OpenAI Five 尚不能击败专业的人类团队）。
2018 年 前列腺癌分级	在对前列腺切除术标本中的前列腺癌进行分级方面，谷歌开发的一种深度学习系统的总体准确度可达到 70%，而据研究，美国委员会认证的一般病理学家的平均准确率仅为 61%。此外，在验证集中，深度学习系统对样本进行分级的准确度已经超过了前 10 位表现最好的病理学家中的 8 位。

六、不足之处

为进一步改进人工智能指数报告，三位人工智能领域专家建议加入新的要素和信息。

（一）常识推理与自然语言理解

IBM 和意大利帕多瓦大学的弗朗西斯卡·罗西（Francesca

¹¹ OpenAI Five 是由 OpenAI 组建的人工智能电竞战队，该系统由五个神经网络组成。

Rossi)指出,深刻的自然语言理解以及常识推理能力仍然是挑战所在,这两项能力的衡量指标有助于推动人工智能系统对话能力以及其他方面的发展。因此,人工智能指数报告中可以添加一些指标来跟踪与这些能力相关的测试和挑战赛的进展,如:威诺格拉德模式挑战赛¹²、AI2的Aristo项目¹³等。此外,如通过人工智能增强人类智能,而非取代人类智能,那么非常重要的一点应是从人机对抗环境向人机合作环境转变。因此,报告中可增加非自主系统数量这一指标,以显示人机合作方面的进展。

(二) 带有人工智能组件的机器人

麻省理工学院的罗德尼·布鲁克斯(Rodney Brooks)希望进一步改善机器人出货量指标。机器人出货量指标的有效性取决于信息源,然而许多信息源提供的工业机器人出货量中仅含有极少的(甚至根本没有)人工智能组件,这就使得机器人出货量指标难以衡量人工智能的真实进展。此外,也可着眼于带有人工智能组件的机器人,如无人机、家庭机器人等。

(三) 政府支出与军事应用

悉尼威尔士大学和柏林工业大学的托比·沃尔什(Toby Walsh)指出,近两年,人工智能领域的最大进展之一可能是政府和行业投资的增加。他建议:一方面增加衡量投资增长的量化指标,通

¹²图灵测试的一个变种,旨在判定AI系统的常识推理能力。

¹³一个能够解答标准化考试问题的人工智能系统。

过相关指标对各国投资进行有效分析，可以衡量人工智能研发重点的转移情况。另一方面，人工智能的军事应用备受关注，可增加可衡量各国军方对人工智能技术的开发和采用情况的指标。

译自：*The AI Index 2018 Annual Report, December 2018 by MIT, Stanford University, OpenAI*

译文作者：工业和信息化部赛迪研究院 魏强 侯雪 陆平 曹茜芮
联系方式：15201609833
电子邮件：weiqiang@ccidthinktank.com

赛迪智库

面向政府 服务决策

思想从这里升华

《赛迪专报》

《安全产业研究》

《产业政策研究》

《赛迪前瞻》

《工业经济研究》

《军民结合研究》

《赛迪智库·案例》

《财经研究》

《工业和信息化研究》

《赛迪智库·数据》

《信息化与软件产业研究》

《科技与标准研究》

《赛迪智库·软科学》

《电子信息研究》

《无线电管理研究》

《赛迪译丛》

《网络安全研究》

《节能与环保研究》

《工业新词话》

《材料工业研究》

《世界工业研究》

《政策法规研究》

《消费品工业“三品”战略专刊》

《中小企业研究》

编辑部：工业和信息化部赛迪研究院

通讯地址：北京市海淀区万寿路27号院8号楼12层

邮政编码：100846

联系人：王乐

联系电话：010-68200552 13701083941

传真：010-68209616

网址：www.ccidwise.com

电子邮件：wangle@ccidgroup.com

报：部领导

**送：部机关各司局，各地方工业和信息化主管部门，
相关部门及研究单位，相关行业协会**

编辑部：工业和信息化部赛迪研究院

通讯地址：北京市海淀区紫竹院路 66 号赛迪大厦 15 层国际合作处

邮政编码：100048

联系人：姚丹

联系电话：（010）88559684 13811086893

传 真：（010）88558833

网 址：www.ccidgroup.com

电子邮件：yaodan@ccidgroup.com

