artificial intelligence index

2018 annual report

指导委员会

Yoav Shoham (*主席*)
斯坦福大学

Raymond Perrault
斯坦福国际研究院

Erik Brynjolfsson
麻省理工学院

Jack Clark
OpenAI

James Manyika
麦肯锡全球研究院

Juan Carlos Niebles
斯坦福大学

Terah Lyons
Partnership On AI

John Etchemendy
斯坦福大学

Barbara Grosz
哈佛大学

项目主管

Zoe Bauer

欢迎阅读人工智能指数2018年度报告

我们的使命是用数据设立探讨人工智能的基础

人工智能指数旨在跟踪、梳理、提取和可视化与人工智能相关的数据。希望本报告能够为政策制定者、研究者、管理人员、记者和公众提供一个全面的数据和分析资源，帮助他们了解复杂的人工智能领域。经“人工智能指数”项目委员会授权，字节跳动联合机器之心对本报告做了中文翻译。

报告原文链接：
http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf

如何引用本报告：

Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John Etchemendy, Barbara Grosz and Zoe Bauer, "The AI Index 2018 Annual Report”, AI Index Committee of the One Hundred Year Study on Artificial Intelligence (AI100), Stanford University, Stanford, CA, December 2018.

(c) 2018 by Stanford University, “The AI Index 2018 Annual Report” is made available under a Creative Commons Attribution-NoDerivatives 4.0 License (International) https://creativecommons.org/licenses/by-nd/4.0/legalcode

目录

人工智能指数2018年度报告简介 5
概述 6
活动规模 8
研究
论文发表 9
课程注册 21
参与 26
机器人软件 29
产业
创业/投资 31
就业 33
专利 35
AI 采用情况调查 36
营收电话会议 40
机器人安装情况 41
开源软件
GitHub 项目统计数据 42
公众兴趣
媒体报道的情绪 43
政府提及情况 44
技术性能
视觉 47
自然语言理解 50
其它指标
衍生指标 55
政府计划 57
达到了人类水平的里程碑 59
有何缺失？ 63
致谢 66
附录 69

人工智能指数2018年度报告简介

我们很高兴推出人工智能指数2018年度报告。本年度的报告要实现两个目标：更新上一年的指标以及尽可能地提供全球性的背景。前一个目标是人工智能指数报告至关重要的使命——为探讨人工智能设定基础，即意味着要持续不断地跟踪规模和技术的进展。但后一个目标也很重要。没有全球化的视野，不可能讲好 AI 故事。2017 年的报告严重偏向了北美地区的活动。这只是因为我们的全球合作关系还有限，并不是因为我们有内在的偏见。今年，我们开始填补这一全球性漏洞。我们认识到，要使本报告真正完备全面，前方还有很长的路要走——会涉及到更多合作和外部参与。

尽管如此，我们仍可以断言 AI 是全球性的。2017 年，Scopus 上 83% 的 AI 论文都来自美国之外。其中来自欧洲的占到了最大的比例（28%）。在全世界范围内，AI 和机器学习（ML）大学课程的入学人数都在增多，尤其值得一提的是中国的清华大学，该校 2017 年人工智能+机器学习组合课程的入学人数比 2010 年时多 16 倍。2017 年，中国生产的工业机器人的数量也占到了世界的最大比重——占全球机器人总出货量的 30%。而且不只是美国、中国和欧洲，其它地方也在推进 AI 的发展。根据我们的统计 AI 专利的方法，韩国和日本在 2014 年产出的 AI 专利分列全球第二和第三，仅次于美国。另外，南非成功举办了第二届 Deep Learning Indaba 大会，这是世界最大型的机器学习教学活动之一，吸引了来自 20 多个非洲国家的超过 500 名参会者。

AI 的多样性不仅体现在地理上。如今，Partnership on AI 已有大约 50 个成员组织是非营利性的，包括美国公民自由联盟、国际特赦组织、牛津大学人类未来研究所和联合国开发计划署。而且人们也对 AI 领域中的性别和少数族裔多样性的重要性有了更高的认识。AI4ALL 和 Women in Machine Learning (WiML)等组织正在加大力度鼓励和支持未得到充分代表的群体参与进来。

下面的指标也能体现本报告所传达的关键信息：AI 是全球性的。

人工智能指数报告概述

本报告分为四大主要部分：

（1）数据：活动规模，技术性能；（2）其它指标：近期的政府计划，衍生指标和人类水平的性能指标；（3）讨论：缺少什么以及参与方式；（4）附录

数据

活动规模指标描述了学术界、公司企业和公众的 AI 活动参与情况。具体的规模数据涵盖从学习 AI 的本科生人数到申请 AI 工作的女性比例，再到AI创业公司风险投资的增长情况。

技术性能指标描述了 AI 性能随时间的变化情况。比如，我们度量了机器问答的质量以及计算机目标检测的训练速度。

在去年的机器人安装和 AI 会议参与情况等指标之外，今年的人工智能指数还添加了国家层面的指标。此外，我们还添加了一些新的指标/研究领域，比如专利、机器人操作系统下载情况、GLUE 指标和 COCO 排行榜。整体而言，我们能看到去年的主要趋势仍在延续：几乎所有领域的 AI 活动都在增长，技术性能也在全面提升。尽管如此，今年还是有一些让我们感到惊讶的情况——比如自然语言方面的巨大进展以及课堂上性别多样性的缺乏。

其它指标

和去年一样，**衍生指标**部分调查了趋势之间的关系。我们还展示了一个探索指标——AI Vibrancy Index（人工智能活力指数）。该指数是将学术界和工业界的趋势结合起来量化 AI 领域的活力。

今年我们还有一个新的定性指标：**近期政府计划**。我们用一页内容简单概括了近期各国政府在人工智能方面的投资和/或陈述。我们囊括了美国、中国和欧洲的计划。随着我们获得的资源的增长以及与新的组织建立合作关系，人工智能指数未来的报告还将包含更多政府数据和分析。

我们的**人类表现里程碑**基于 AI 展现出人类水平或超人类水平能力的实例的时间线。我们加入了 2018 年的 4 项新成就

人工智能指数报告概述（续）

有何缺失？部分汇集了该领域的专家的反馈——哪些指标有助于更全面和更有代表性的人工智能指数。我们对有何缺失？的贡献者表示感谢，他们还会为未来报告的内容补充提供信息。

附录

附录部分能为读者提供完全透明的资源、方法和微妙之处的描述。附录中还包含了本报告中几乎每张图的基础数据。我们希望 AI 社区的每位成员都能与他们认为最能代表该领域的数据和指标互动。

符号

当我们展示 AI 的广度数据时，我们都会在该页面上标记一个代表全球的符号。这类数据包括国家比较、美国之外地区的情况以及有关 AI 社区多样性的数据。

活动规模

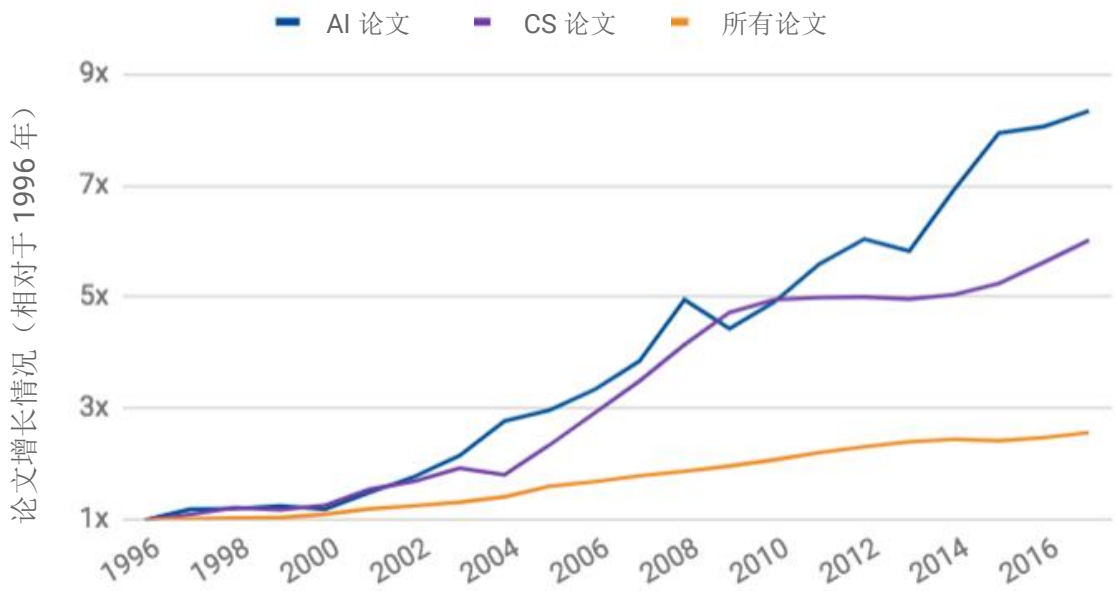
论文发表情况：按主题统计

下表列出了相对于 1996 年的学术论文发表率的发表率年度增长情况。这幅图比较了所有领域、计算机科学（CS）领域和人工智能（AI）领域的论文增长情况。

AI 领域年度论文发表数量的增长持续快于 CS 领域年度论文发表数量的增长，这说明 AI 论文发表数量的增长动力不仅仅是研究者对计算机科学的兴趣的增长。

参见 附录 1 了解相关数据和方法。

按主题统计的论文发表年度增长情况（1996-2017）
来源：Scopus



注：这幅图使用了 Scopus 查询搜索关键词"Artificial Intelligence"，而不是 Elsevier 关键词方法。参阅附录了解更多细节。

AI 发展快于计算机科学

自 1996 年以来，Scopus 上的 AI 论文数量增长了 8 倍。CS 论文数量在同一段时间里增长了 6 倍。

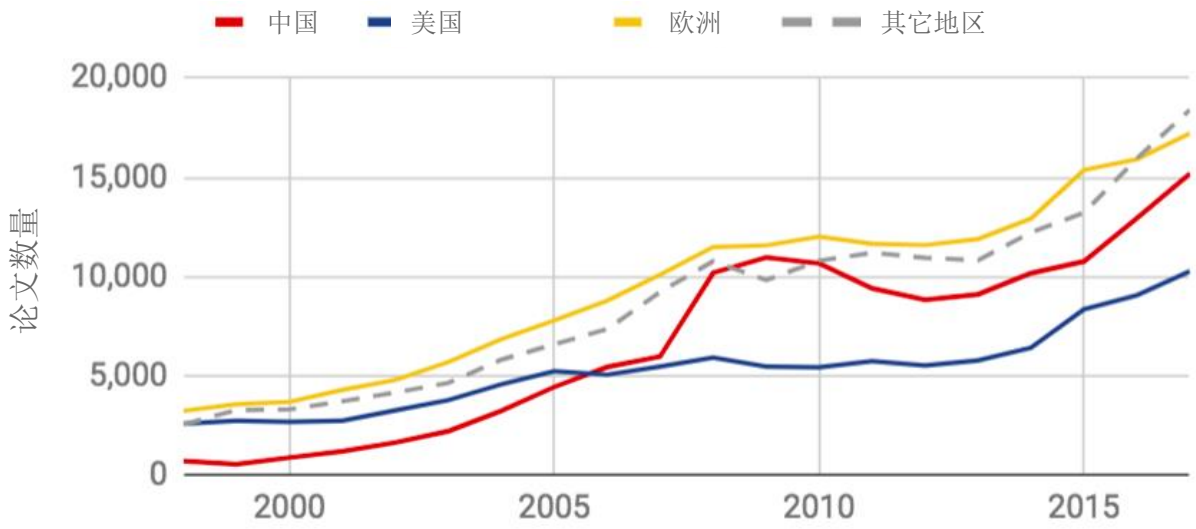
论文发表情况：按地区统计

下图展现了不同地区AI论文年度发表数量。

欧洲一直以来都是发表 AI 论文最多的地区，2017 年 Scopus 上有 28% 的 AI 论文源自欧洲。与此同时，从 2007 年到 2017 年，中国发表的 AI 论文数量增长了 150%。但是，中国论文数量在 2008 年左右经历了一次猛增之后又出现了一段时间的回落。

参见 附录 2 了解相关数据和方法。

Scopus 上按地区统计的 AI 论文年度发表情况（1998-2017）
来源：Elsevier



注：我们推测中国在 2008 年 AI 论文数量增长的原因是 国家中长期科学和技术发展规划（2006-2020）以及其它一些政府项目为 AI 研究提供了资金和一系列激励政策。类似地，FP7 (2007–2013) 等欧洲的科学和技术研究规划也可能贡献了 2008-2010 年欧洲论文数量的少量增长。

欧洲是发表 AI 论文最多的地区

2017 年，Scopus 上 28% 的 AI 论文与欧洲的作者有关，其后是中国（25%）和美国（17%）。

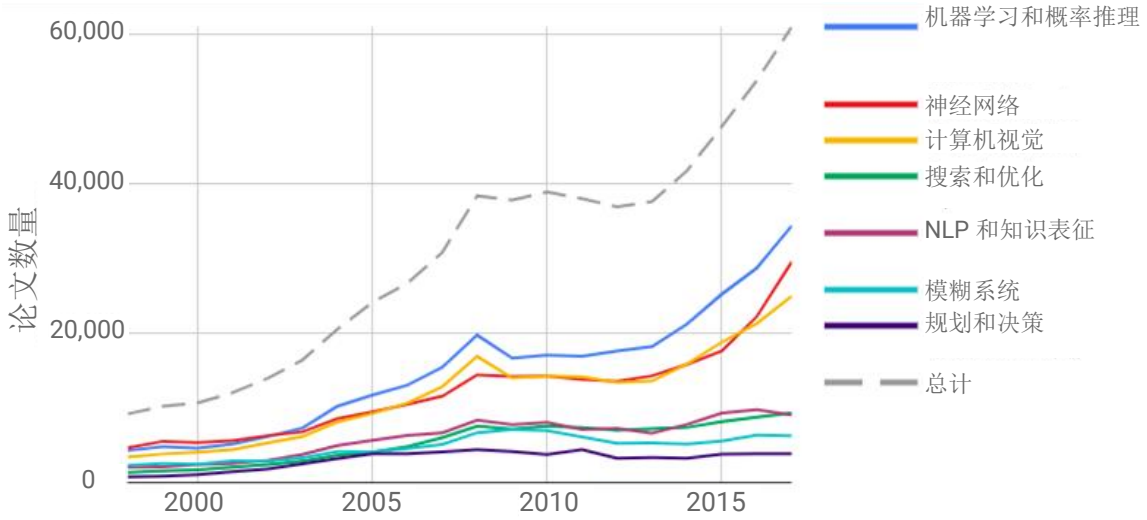
论文发表情况：按子类别统计

按子类别统计的 Scopus 上的 AI 论文数量。这些子类别并不互斥。

2017 年，56% 的论文属于机器学习和概率推理类别；相比而言 2010 年时这一数字为 28%。 对于下面的大多数类别，2014-2017 年之间的论文发表速度比 2010-2014 年之间的速度都更快。最值得注意的是，神经网络类别的论文数量在 2010-2014 年间的复合年增长率（CAGR）为 3%，在 2014-2017 年间为 37%。

参见 附录 2 了解相关数据和方法。

按子类别统计的 Scopus 上的 AI 论文数量（1998—2017）
来源：Elsevier



Scopus 上的神经网络论文数量在 2014-2017 年之间的复合年增长率为 37%。

论文发表情况：arXiv 上的 AI 论文

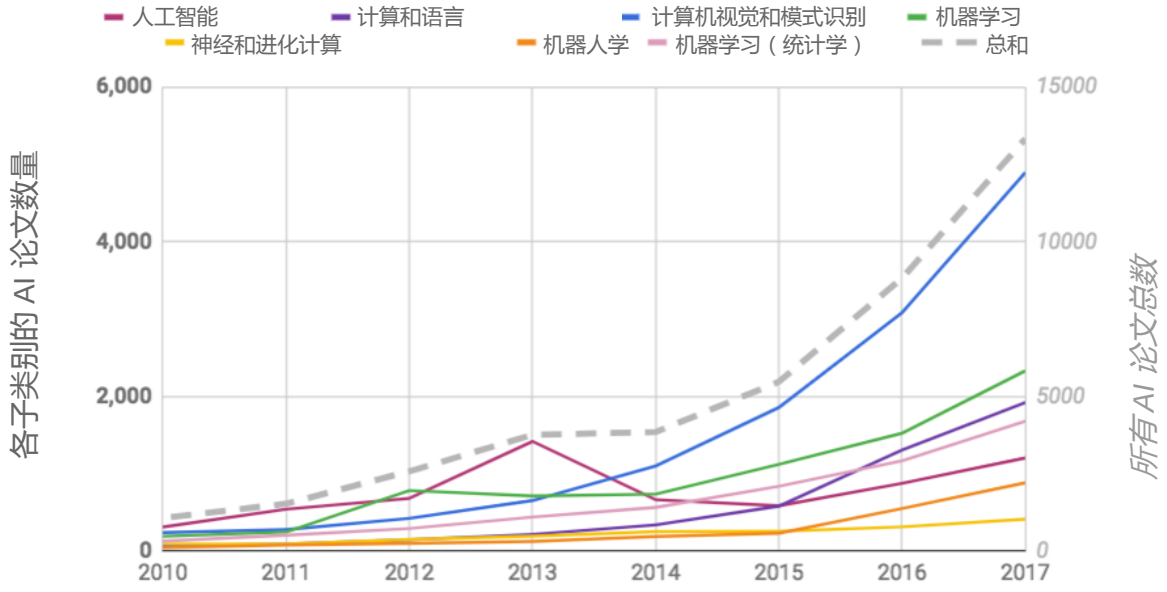
按每篇论文的主要子类别统计的 arXiv 上 AI 论文的数量。右轴的标度是 arXiv 上所有 AI 论文的总数。论文的数量是累积的。

arXiv 上 AI 论文的整体数量以及某些子类别的论文数量在增长。这说明 AI 论文的作者倾向于传播他们的研究，无论是否经过同行评议或被著名的 AI 会议接收。这还表明了该领域的竞争性。

“计算机视觉和模式识别”自 2014 年以来一直都是 arXiv 上最大的 AI 子类；在 2014 年之前，这个类别的论文增长紧跟“人工智能”和“机器学习”。这说明研究者对计算机视觉（及其一般应用）的兴趣正在增长，另外数据还表明“计算和语言”和“机器人学”等其它 AI 语言领域也在增长。

参见 附录 3 了解相关数据和方法。

arXiv 上各子类别的 AI 论文数量（2010-2017）
来源：arXiv



“……除了出版物的增长之外，需要重点注意这些社区正采用 arXiv 来传播研究结果。我们已经见过很多次如何建立某些关键团体，然后再在社区内进一步催生更高层面的参与……”
Paul Ginsparg，康奈尔大学

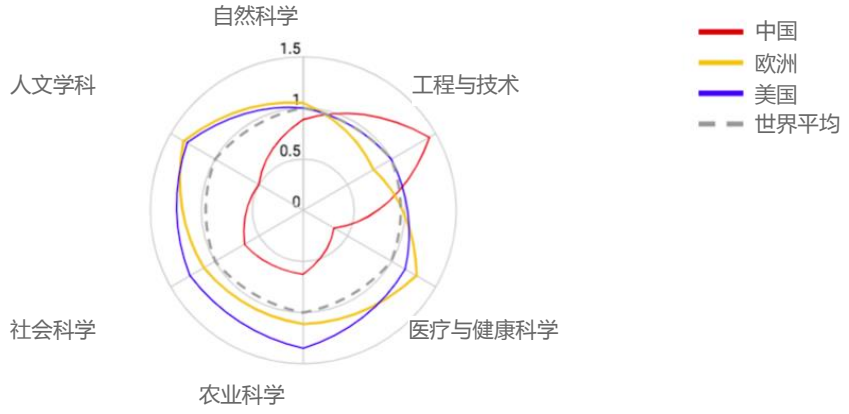
论文发表情况：不同地区的关注重点

相对活动指数（RAI/Relative Activity Index）是通过比较地区研究活动与全球研究活动来近似了解一个地区的专业化程度。RAI 的定义是一个国家的出版物产出在 AI 领域的份额相比于全球的出版物产出在 AI 领域的份额。RAI 的值为 1.0 时表示该国在 AI 领域的研究活动正好与全球范围的 AI 研究活动一致。其值大于 1.0 时表示该国更重视 AI，其值小于 1.0 则表示该国更不重视 AI。

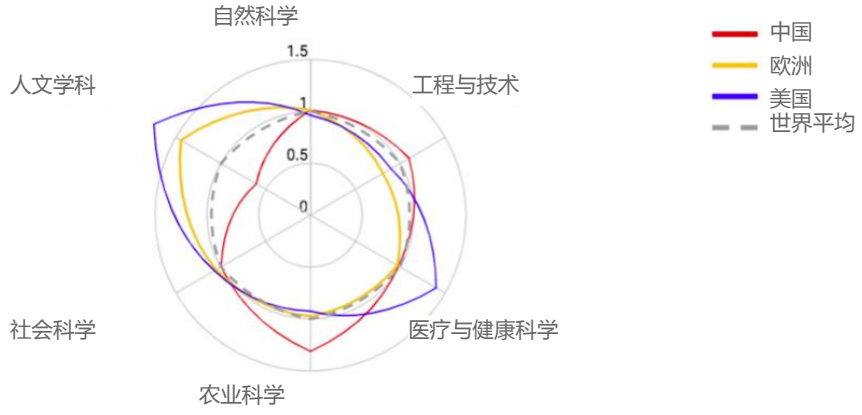
中国更重视“工程与技术”和“农业科学”，而美国和欧洲则更重视“人文学科”和“医疗与健康科学”。相比于 2000 年的数据，2017 年的数据表明这三个地区的专业化程度都在提升，中国对农业的重视大幅提升。这符合我们的预期，因为中国是最大的食品生产国，而且倾向于研究 AI 的应用。

参见 附录 2 了解相关数据和方法。

按地区和 AI 研究所涉领域统计的相对活动重心（2000）
来源：Elsevier



按地区和 AI 研究所涉领域统计的相对活动重心（2017）
来源：Elsevier



论文发表情况：论文所属机构

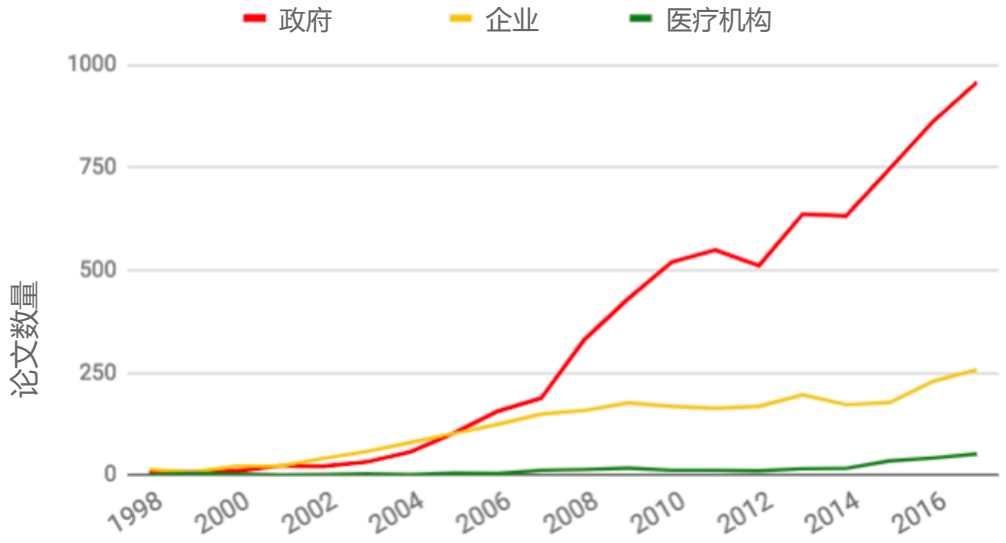
来自政府、企业和医疗机构的 Scopus 论文数量。以下是中国的数据。欧洲和美国的数据在下一页。第 16 页直接在一张图中比较了这三个地区。我们展示了一张企业论文增长的图片以及一张政府论文增长的图片。2017 年，中国政府产出的 AI 论文比中国企业多近 4 倍。自 2007 年以来，中国源自政府的 AI 论文增长了 400%，同期企业 AI 论文仅增长了 73%。

相比于中国和欧洲，美国的企业 AI 论文占比更大。2017 年，美国企业 AI 论文的比例比中国企业 AI 论文的比例高 6.6 倍，比欧洲的高 4.1 倍。

注意这三个地区的学术论文（未展示）都远多于政府、企业和医疗机构的论文。参阅每张图下的注释。

参见 附录 2 了解相关数据和方法。

来自不同机构的论文——中国（1998-2017）
来源：Elsevier



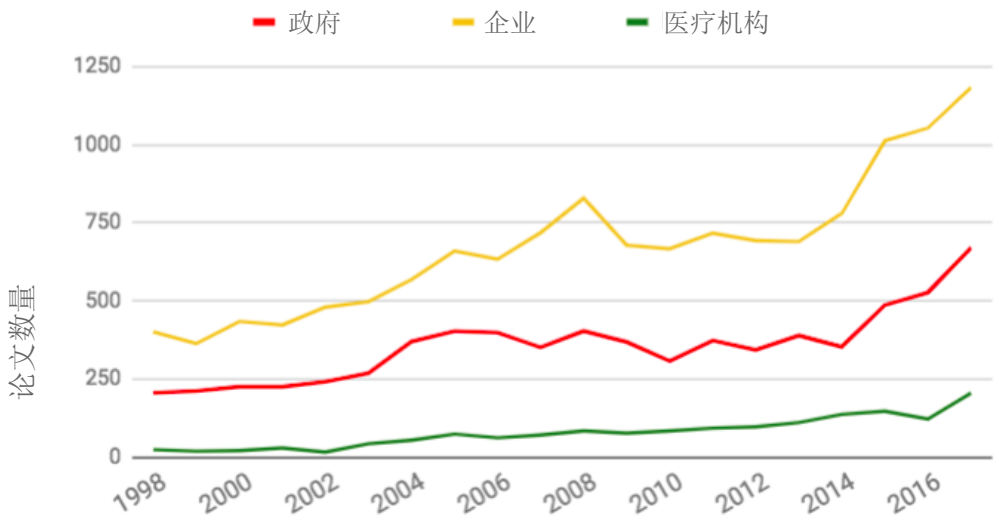
注：参考，Scopus 上 2017 年来自中国学术界的 AI 论文的数量（图中未给出）大约为 14 000，约占中国 AI 论文总数的 92%。

中国的政府 AI 论文增长了 400%

从 2007–2017 年，中国的政府 AI 论文增长了 400%，而企业 AI 论文仅增长了 73%。

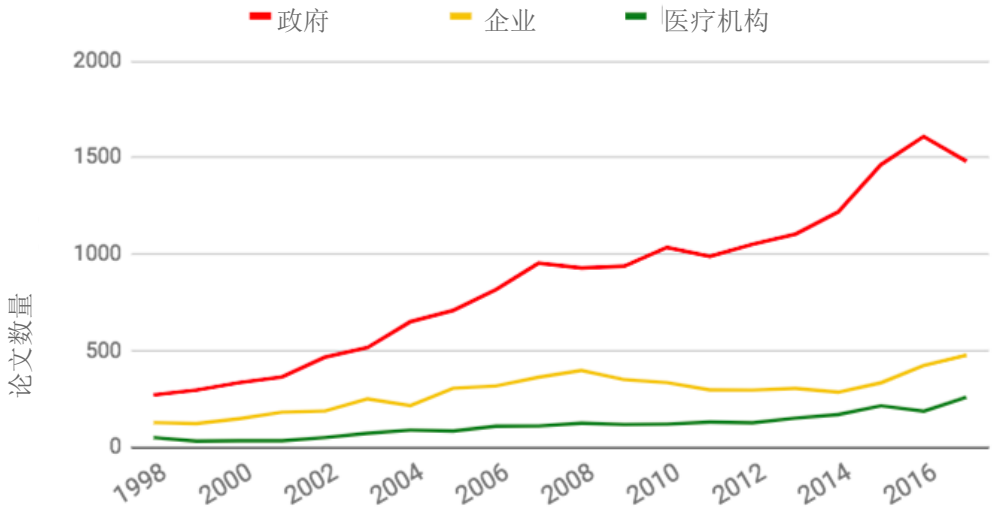
论文发表情况：论文所属机构（续）

来自不同机构的论文——美国（1998-2017）
来源：Elsevier



注：参考，Scopus 上 2017 年来自美国学术界的 AI 论文的数量（图中未给出）大约为 9 000，约占美国 AI 论文总数的 81%。

来自不同机构的论文——欧洲（1998-2017）
来源：Elsevier



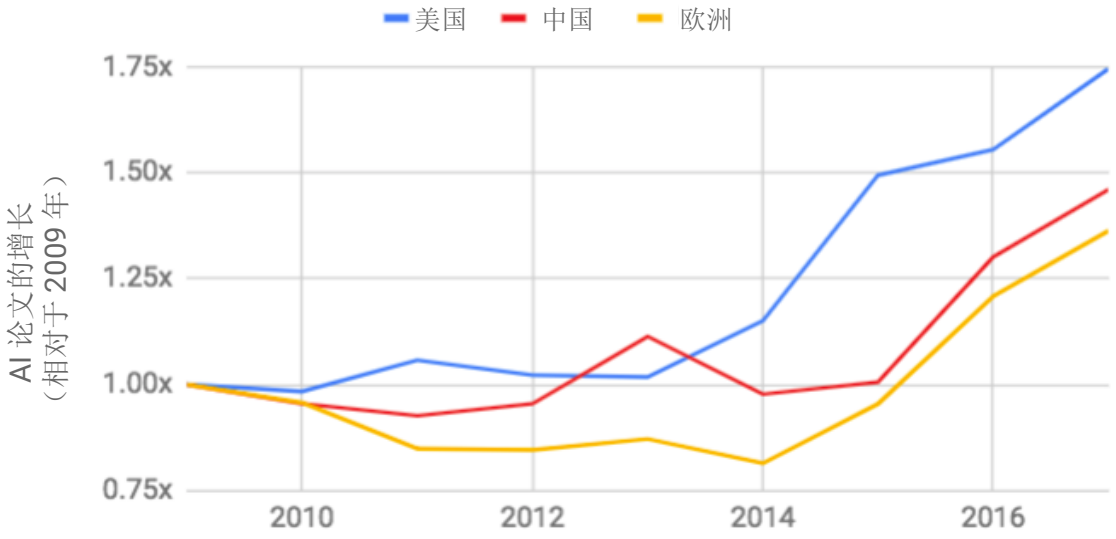
注：参考，Scopus 上 2017 年来自欧洲学术界的 AI 论文的数量（图中未给出）大约为 16 000，约占欧洲 AI 论文总数的 88%。

美国的企业论文多 6.6 倍

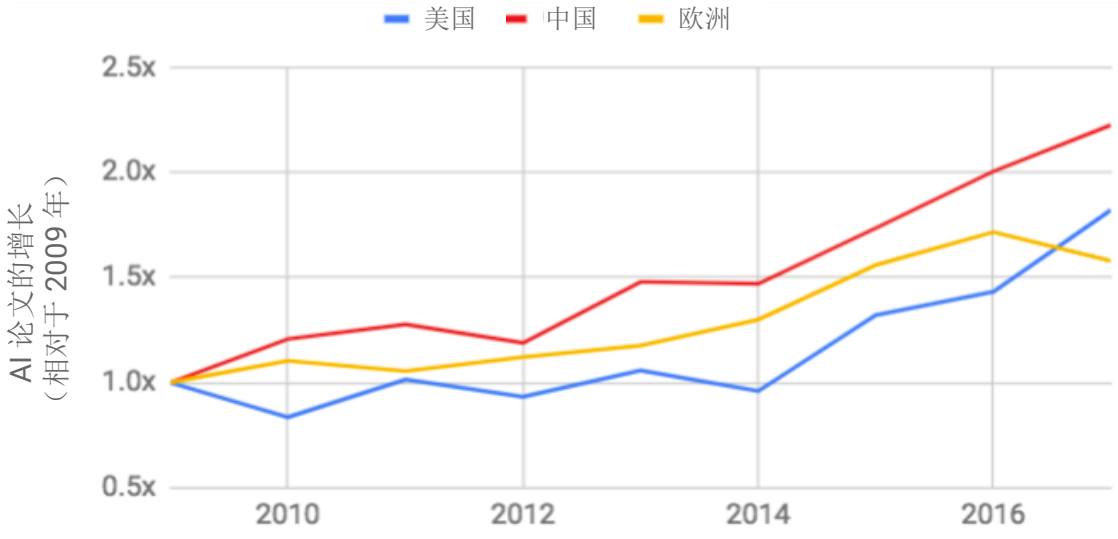
2017 年，美国企业 AI 论文的比例比中国企业 AI 论文的比例高 6.6 倍，比欧洲的高 4.1 倍。

论文发表情况：论文所属机构（续）

企业 AI 论文增长情况（2009-2017）
来源：Elsevier



政府 AI 论文增长情况（2009-2017）
来源：Elsevier



中国的政府 AI 论文数量自 2009 年以来已经翻番

与此同时，美国的企业 AI 论文增长最显著。2017 年的企业 AI 论文数量是 2009 年时的 1.7 倍。

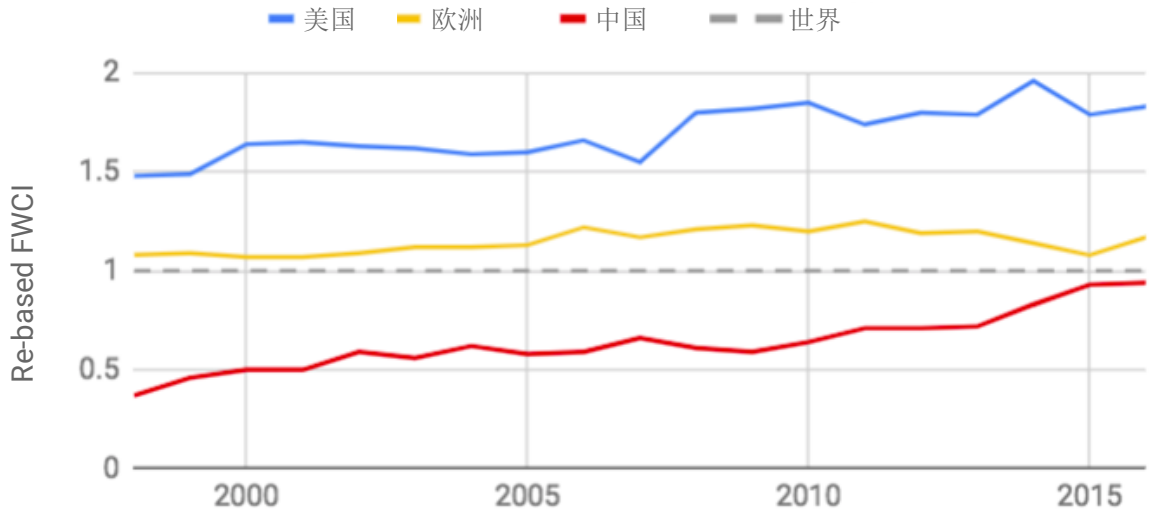
论文发表情况：按地区统计的引用影响

下图展示了按地区统计的 AI 作者的平均领域加权的引用影响。领域加权的引用影响（FWCI）是作者在该领域得到的引用平均数除以所有 AI 作者的引用平均数。在这张图中，显示的 FWCI 基准进行过调整（re-based FWCI），也就是说展示的是相对于世界平均值的被引用影响。 re-based FWCI 为 1 时表示出版情况与世界平均一致。re-based FWCI 为 0.85 时表示引用量比世界平均低 15%。

尽管欧洲的年度 AI 论文发表数量最大，但欧洲的 re-based FWCI 相对平稳，且与世界平均相近。相对而言，中国的 re-based FWCI 有明显增长。中国的 AI作者在 2016 年的被引用量比 2000 年时多 44%。美国在 re-based FWCI 上的表现持续优于其它地方。美国作者的被引用量比全球平均水平高 83%。

参见 附录 2 了解相关数据和方法。

按地区统计的领域加权的引用影响（1998—2017）
来源：Elsevier



与 2000 年时相比，中国的 AI 作者的被引用量增长了 44%。

论文发表情况：论文作者在不同地区之间的迁移

下一页的图片展示了 AI 作者的跨国流动对论文生产和引用的影响。

我们调查了 4 种不同的流动类别：固定作者、临时流动作者、迁移性流入作者和迁移性流出作者。**固定作者**是指从未在他们的国家地区之外发表过论文的活跃研究者。**临时流动作者**是指在他们的国家地区之外、在两年或两年以内时间里发表文章的研究者。**迁移性作者**是指在他们的国家地区之外、在超过两年时间里发表文章的研究者。 作者属于**迁移性流出**还是**迁移性流入**取决于图表。

X 轴是相对发表率——每个类别中作者的平均论文数量除以该地区总体的平均论文数量。Y 轴是领域加权的引用影响（FWCI）——每个流动类别中作者得到的平均引用数除以所有地区总体的平均引用数。

这个分析仅考虑了 AI 作者。如果一位研究者至少有 30% 的论文涵盖了人工智能，则就认为他或她是 AI 作者。作者的国家地区是指他们发表的第一篇论文的地区。

…

在美国、中国和欧洲，发布率最低的都是固定作者。此外，在这三个地区中，迁移性作者（包括迁移性流入和流出）都有最高的 FWCI。因此，流动的作者往往有更高的引用数量且发表论文的频率也往往更高。

在这三个地区中，中国的固定 AI 作者所占的比重最大（76%），之后是欧洲（52%），然后是美国（38%）。尽管更大比例的中国作者是固定作者，但中国的非固定作者比其它两个地区的非固定作者有更高的发表率。换句话说，尽管地理上流动的中国作者相对较少，但他们往往比其它地方的流动性作者平均更高产。

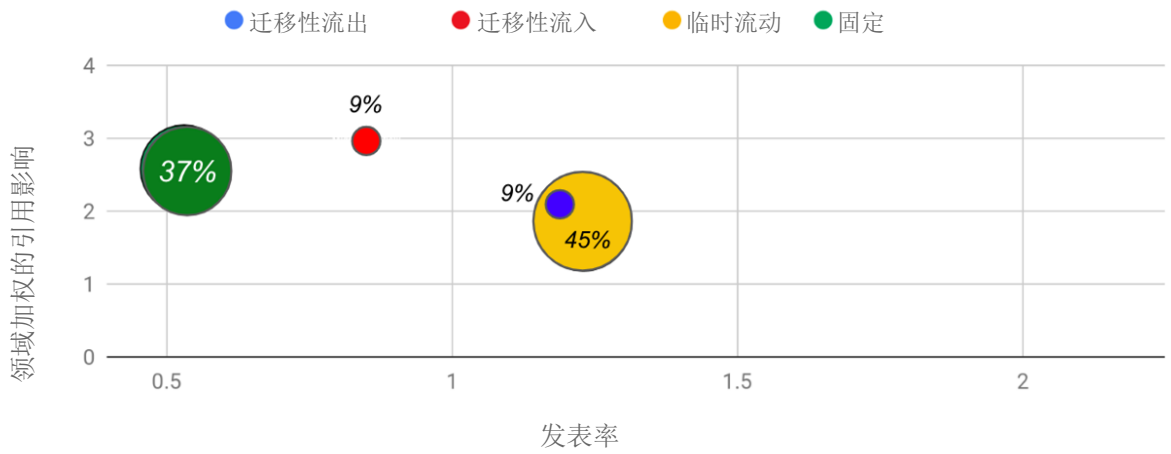
参见 附录 2 了解相关数据和方法。

流动的作者有更高的引用数且发论文频率更高。

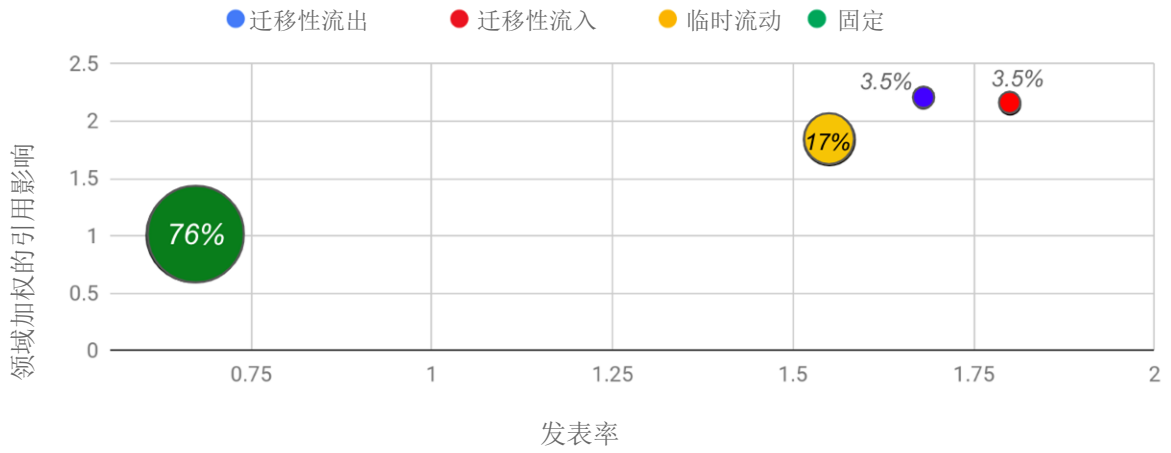
论文发表情况：论文作者在不同地区之间的迁移

描述见前一页。

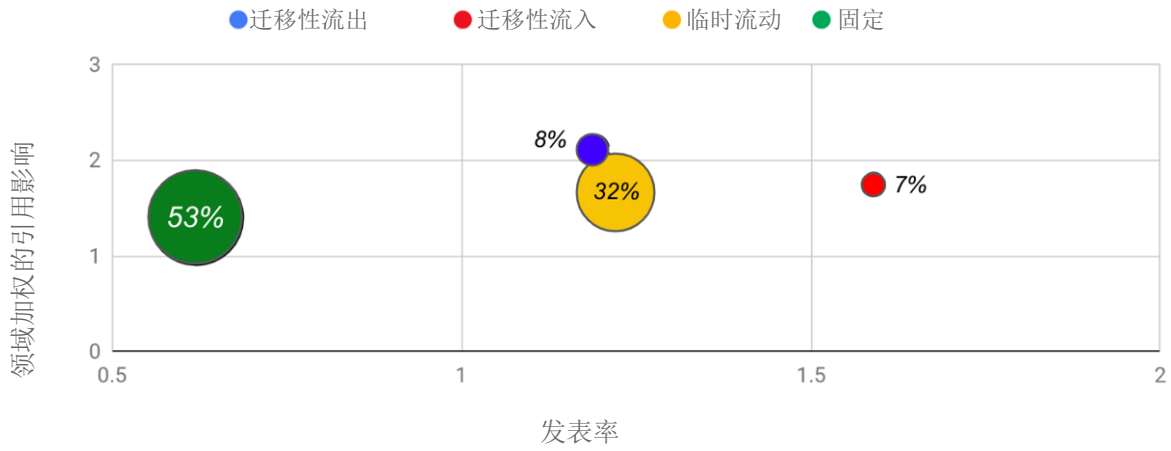
不同的国际流动类别的发表率和引用影响——美国（1998-2017）
来源：Elsevier



不同的国际流动类别的发表率和引用影响——中国（1998-2017）
来源：Elsevier



不同的国际流动类别的发表率和引用影响——欧洲（1998-2017）
来源：Elsevier



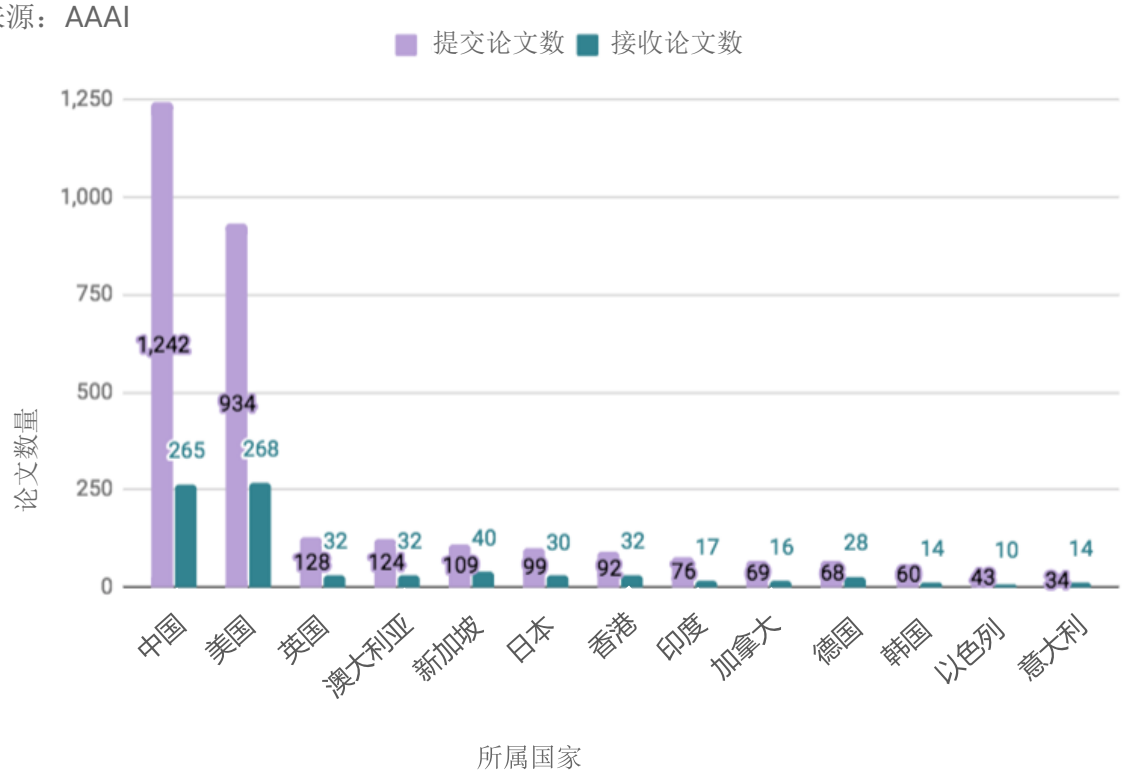
论文发表情况：按国家统计的 AAAI 论文

下图展示了 2018 年不同国家向人工智能促进协会（AAAI）会议提交的论文数和被接收的论文数。AAAI 2018 会议于 2018 年 2 月在路易斯安那州新奥尔良举办。

2018 年提交给 AAAI 的论文中有 70% 来自美国或中国。尽管中国提交的论文数量最大，但美国和中国被接收的论文数很接近，分别为 268 和 265。因此，美国论文的接收率为 29%，而中国论文的接收率为 21%。德国和意大利论文的接收率最高（41%），但他们提交的论文也更少。

参见 附录 4 了解相关数据和方法。
参会人数随时间的变化情况见第 24 页。

提交和接受的论文的数量——AAAI 2018 会议
来源：AAAI



70% 的 AAAI 论文来自美国或中国

2018 年的 AAAI 会议有 70% 的提交论文和 67% 的接收论文来自美国或中国。

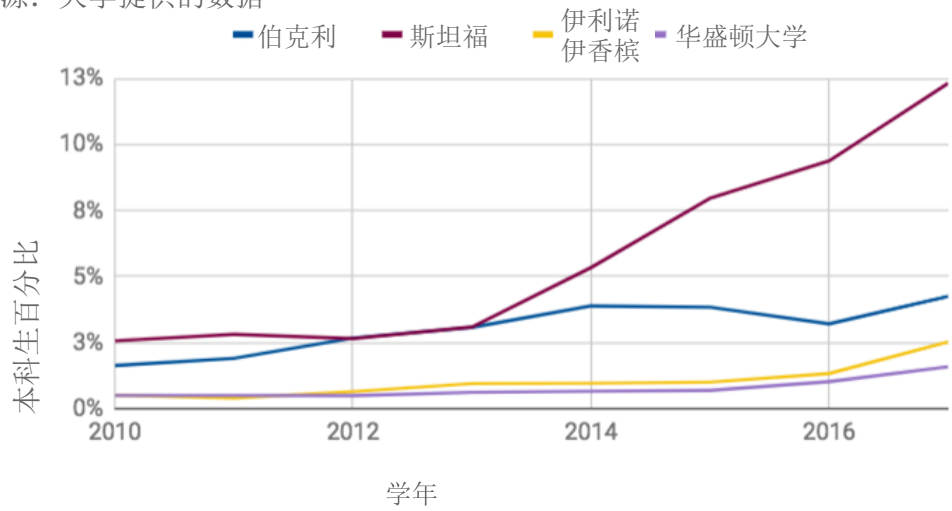
课程注册情况：美国的人工智能课程

下图展示了注册入门级 AI 和 ML 课程的本科生的百分比。这是一个考虑了学校规模后计算得到的指标。学校选择标准、实际注册数量和完整的大学名单可参见附录。

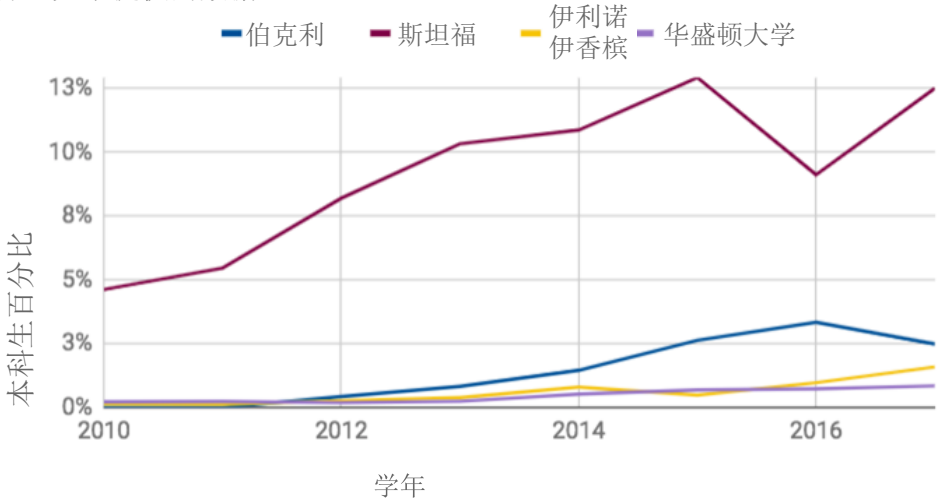
尽管入门级 AI 课程往往比入门级 ML 课程有稍微更大比例的本科生（AI 课程平均占 5.2%，ML 课程平均占 4.4%），但选择入门级 ML 课程的本科生数量的增长速度更快。
增长情况比较见下一页。这说明机器学习作为一个 AI 子领域的重要性正日益增长。

参见 附录 5 了解相关数据和方法。 *完整大学名单见附录。*

注册入门级 AI 课程的本科生比例（2010-2017）
来源：大学提供的数据



注册入门级 ML 课程的本科生比例（2010-2017）
来源：大学提供的数据



有关性别多样性的备注：
斯坦福 2017 年入门级 AI 课程：74.45% 男性（共 869 位学生）；伯克利 2017 年入门级 AI 课程：73.37% 男性（共 973 位学生）
斯坦福 2017 年入门级 ML 课程：75.91% 男性（共 876 位学生）；伯克利 2017 年入门级 ML 课程：78.67% 男性（共 720 位学生）

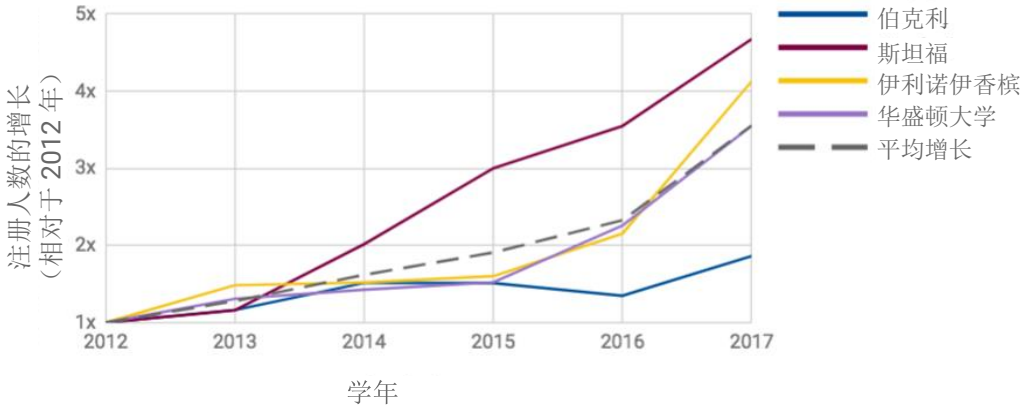
课程注册情况：美国的人工智能课程（续）

下图展示了根据美国的学校选择标准在计算机科学方面领先的一些大学的 AI 和 ML 课程注册情况，实际注册人数和完整大学名单见附录。

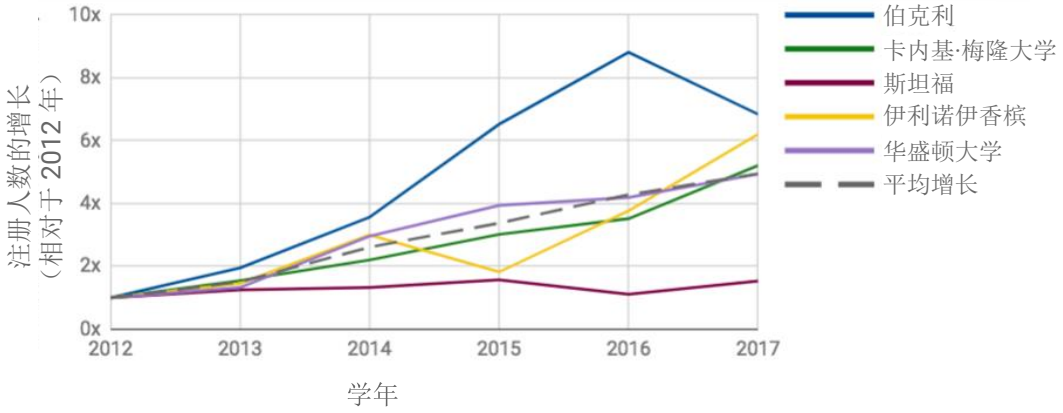
2017 年的入门级 AI 课程注册人比 2012 年多 3.4 倍，2017 年的入门级 ML 课程注册人数比 2012 年时多 5 倍。加利福尼亚大学伯克利分校 2017 年的入门级 ML 课程的学生数量是 2012 年的 6.8 倍。该课程的增长超过我们调研的其它任何入门级 ML 课程。

参见附录 5 了解相关数据和方法。

入门级 AI 课程注册人数的增长（2012—2017）
来源：大学提供的数据



入门级 ML 课程注册人数的增长（2012—2017）
来源：大学提供的数据



2017 年入门级 ML 课程注册人数是 2012 年时的 5 倍

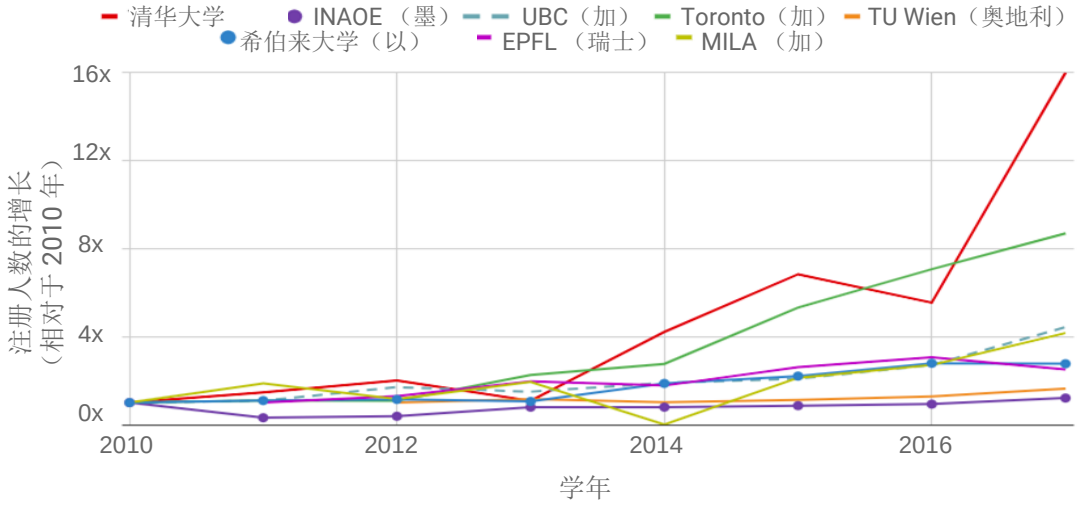
课程注册情况：其它国家的人工智能课程

下面以及下一页的图片展示了美国之外几所计算机科学领先的大学的 AI 和 ML 课程注册情况。第一张图展示了提供了 2010-2017 学年数据的美国之外的学校的相对增长情况。第二张图展示了提供了 2016-2018 学年数据的美国之外的学校的相对增长情况。学校选择标准、实际注册数量和完整的大学名单可参见附录。

清华大学 2017 年的 AI+ML 组合课程的注册人数是 2010 年时的 16 倍。在被调研的学校中，我们发现 AI 课程注册人数的增长相对取决于具体学校，并不会受到地理位置的显著影响。人工智能指数研究有望在未来的报告中改进这一假设。

参见 附录 6 了解相关数据和方法。

AI+ML 课程注册人数的增长——非美国（2010-2017）
来源：大学提供的数据

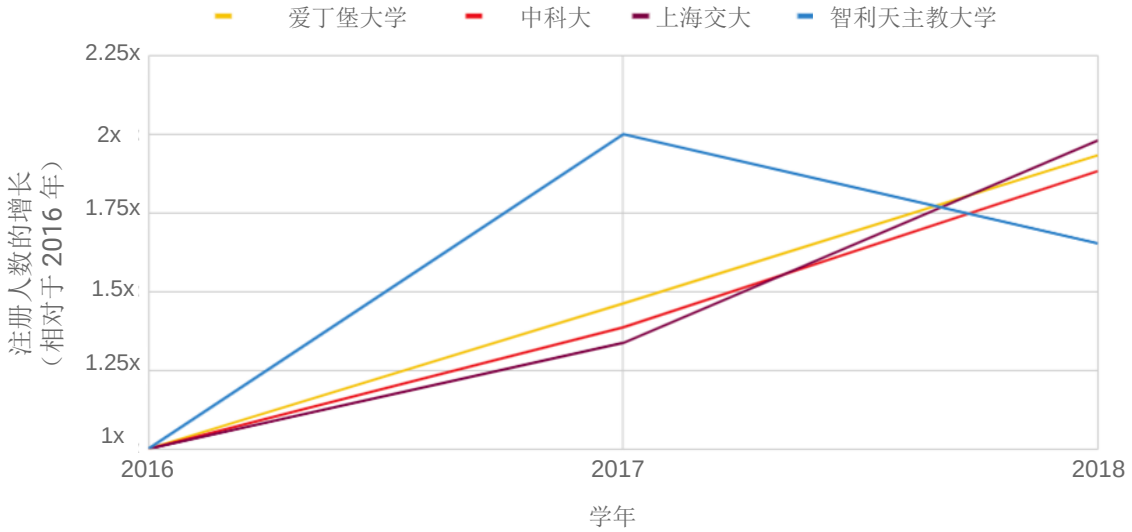


注：这幅图展示了可获得过去学年的数据的大学。所有的增长率都相对于 2010 年的注册人数，但 EPFL（2011 年）和 TU Wien（2012）年除外。

课程注册情况：其它国家的人工智能课程（续）

描述见前一页

AI 课程注册人数的增长——非美国（2016-2018）
来源：大学提供的数据



注：这幅图展示了历史数据有限的大学。所有的增长率都相对于 2016 年的注册人数。

教职人员多样性

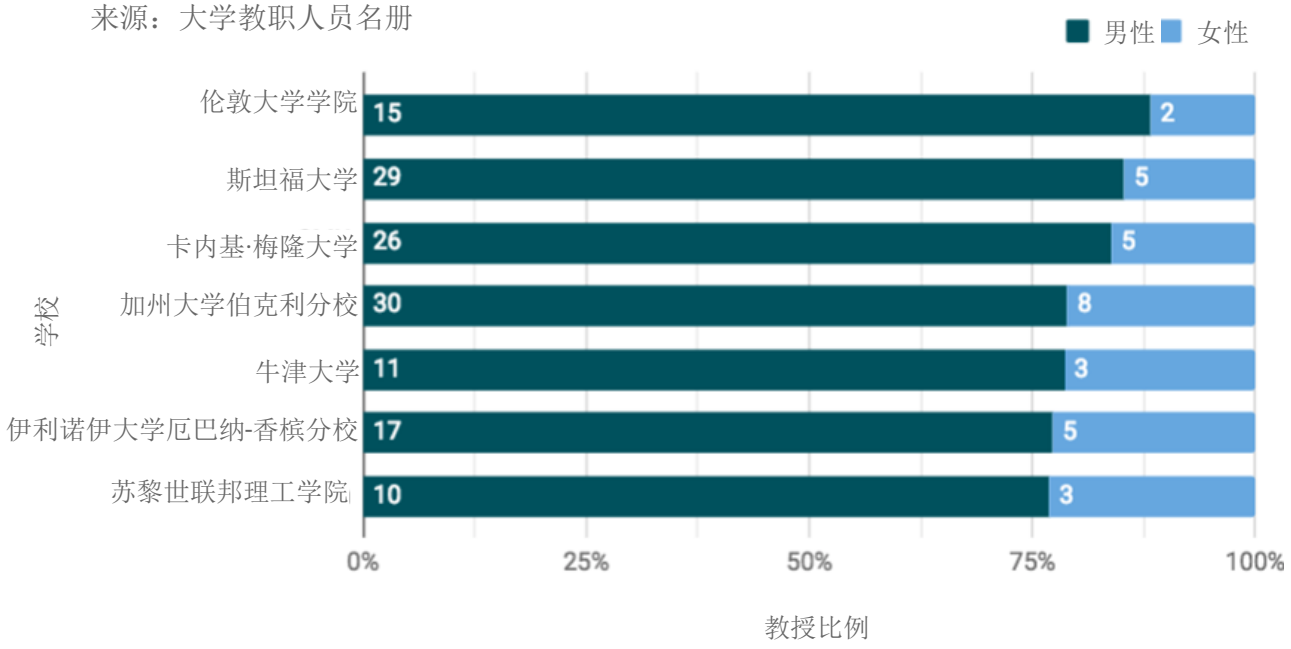
下图展示了几所计算机科学领先的大学的 AI 教授性别分布。数据是在 2018 年 9 月 21 日使用教职人员名册收集的。我们选择了容易获取 AI 教职人员名册的学校。

多样性提升的一大显著障碍是难以获取有关产业界和学术界多样性的统计数据。我们希望组织机构在多样性统计数据上能更加透明，以帮助改善该领域的多样性。

在被调研的学校中，我们发现平均 80% 的 AI 教授都是男性。我们也注意到，不管地理位置如何，不同学校之间存在很小的差异。尽管如此，由于调研的学校数量有限，这些发现仅算得上是冰山一角。

参见 附录 7 了解相关数据和方法。

所选学校的 AI 教授的性别分布（2018 年 9 月）
来源：大学教职人员名册



80% 的 AI 教授是男性。

平均而言，加州大学伯克利分校、斯坦福大学、伊利诺伊大学厄巴纳-香槟分校、卡内基·梅隆大学、伦敦大学学院 、牛津大学和苏黎世联邦理工学院的AI 教授有 80% 是男性

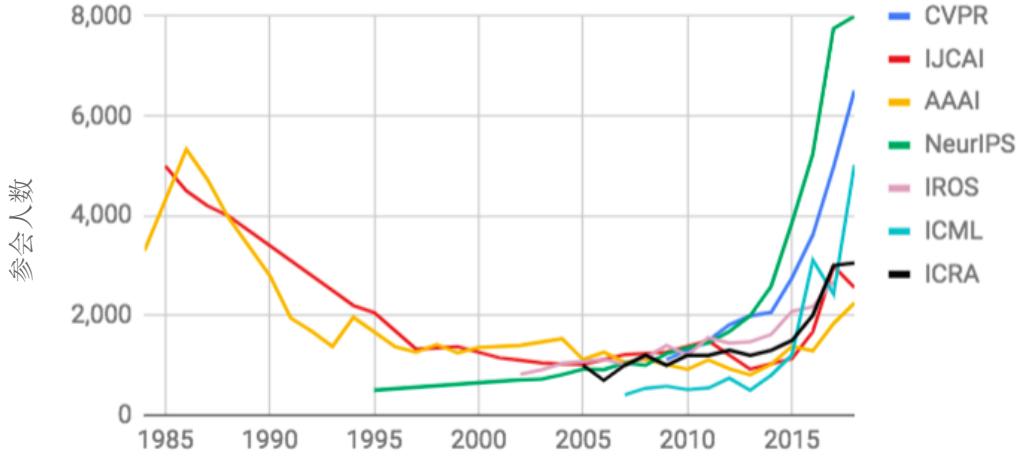
参与：大型人工智能会议

下面的图片展示了大型 AI 会议的参会人数以及大型 AI 会议的参会人数相对于 2012 年的增长情况。大型 AI 会议是指 2017 年有超过 2000 位参会者的会议。

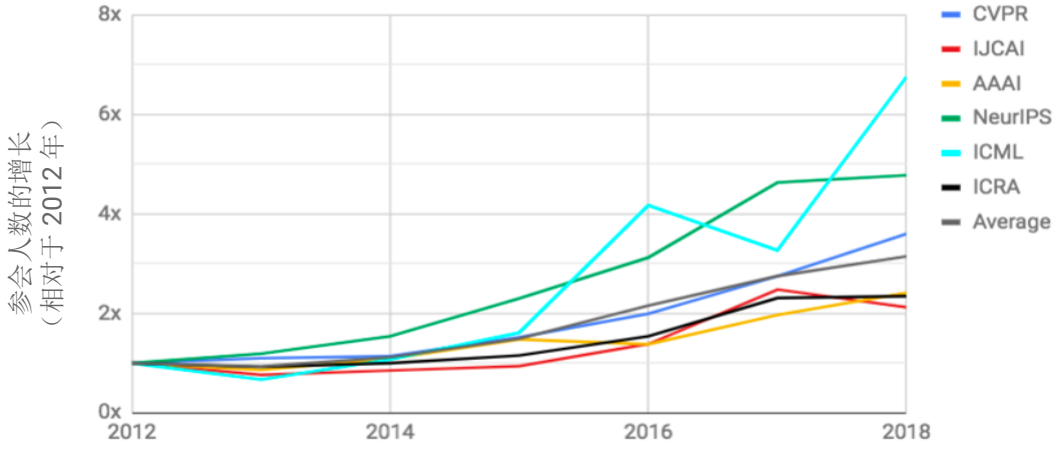
NeurIPS（原 NIPS）、CVPR 和 ICML 是参会人数最多的 AI 会议，这些会议的参会人数自 2012 年以来增长最显著。NeurIPS 和 ICML 的增长速度最快——分别达到了其 2012 年参会人数的 4.8 和 6.8 倍。这表明 ML 作为 AI 的一个子领域，正得到越来越大的关注。与此同时，关注符号推理的会议的增长一直相对较小。

参见 附录 8 了解相关数据和方法。

大型会议的参会人数（1984–2018）
来源：会议提供的数据



大型会议参会人数的增长情况（2012–2018）
来源：会议提供的数据



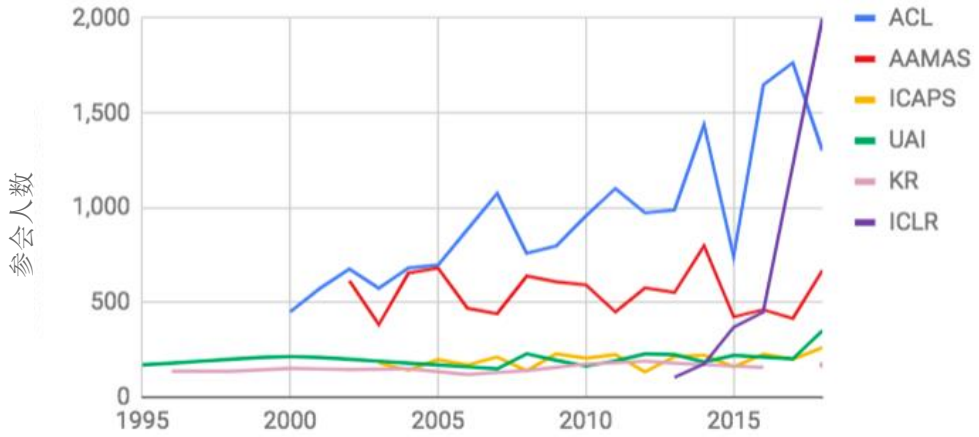
参与：小型人工智能会议

下面的图片展示了小型 AI 会议的参会人数以及小型 AI 会议的参会人数相对于 2012 年的增长情况。小型 AI 会议是指 2017 年参会人数少于 2000 的会议。

ICLR 2018 年的参会人数是 2012 年的 20 倍。这一显著增长的原因很可能是当今 AI 领域更加重视深度学习和强化学习。

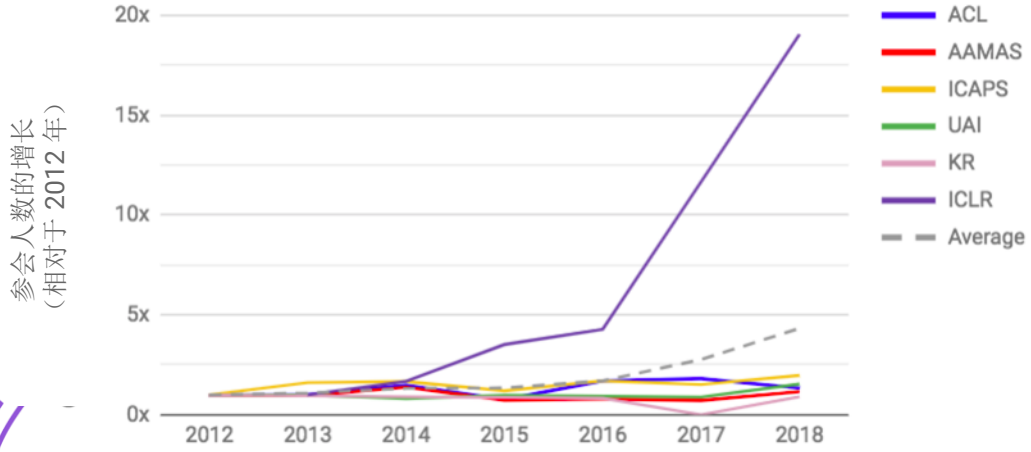
参见附录 8 了解相关数据和方法。

小型会议的参会人数（1984-2018）
来源：会议提供的数据



注： KR 在 2018 年才第一次举办研讨会（某些参加研讨会的人也参加了会议）。为了数据的一致性，上图中 KR 的参会人数没有包含那 106 位研讨会参会者。

小型会议参会人数的增长（2012-2018）
来源：会议提供的数据



ICLR 2018 年的参会人数比 2013 年多 20 倍

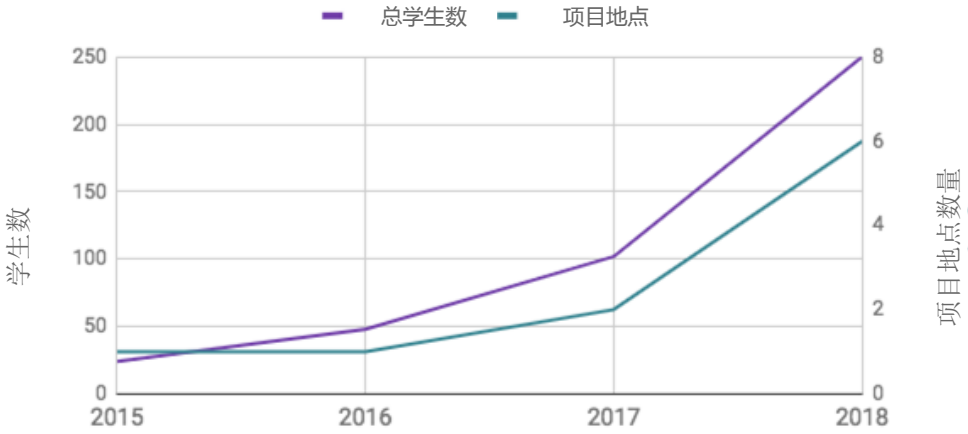
参与：多样化组织机构

下面的图片展示了 Women in Machine Learning（WiML）主办的年度研讨会的注册人数和 AI4ALL 的学生数量。WiML 是一个致力于支持机器学习领域中的女性的组织；AI4ALL 是一个专为增加 AI 领域的多样性和包容性而设计的 AI 教育倡议。

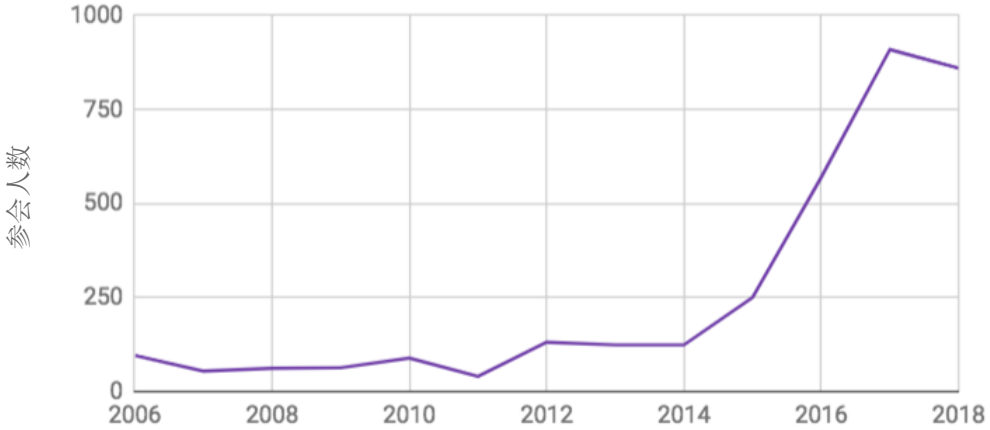
过去几年，WiML研讨会和 AI4ALL 的注册人数都有增长。在 2018 年，WiML 研讨会参会者数量比 2014 年时多 600%，AI4ALL 的学生数比 2015 年时多 900%。这些增长说明 AI 领域在持续不断地努力纳入女性和代表性不足的群体。

参见附录 8 了解相关数据和方法。

AI4ALL 学生和项目数量（2015-2018）
来源：AI4ALL



WiML注册人数（2006-2017）
来源：WiML



注： WiML 注册人数在 2017 年有较大增长，因为 WiML 研讨会的议程由前几年的一天变成了两天。

“……这一数据能反映近几年我们在 WiML 社区所看到的增长，而且我们认为这也与 AI 领域内有意识地实现社区多样化有关。但我们还有很长的路要走，因为参与 WiML 活动的人数仍仅有数百，而 NeurIPS 等通常在同一位置举办的会议的参会人数通常能达到数千……”
WiML 理事会

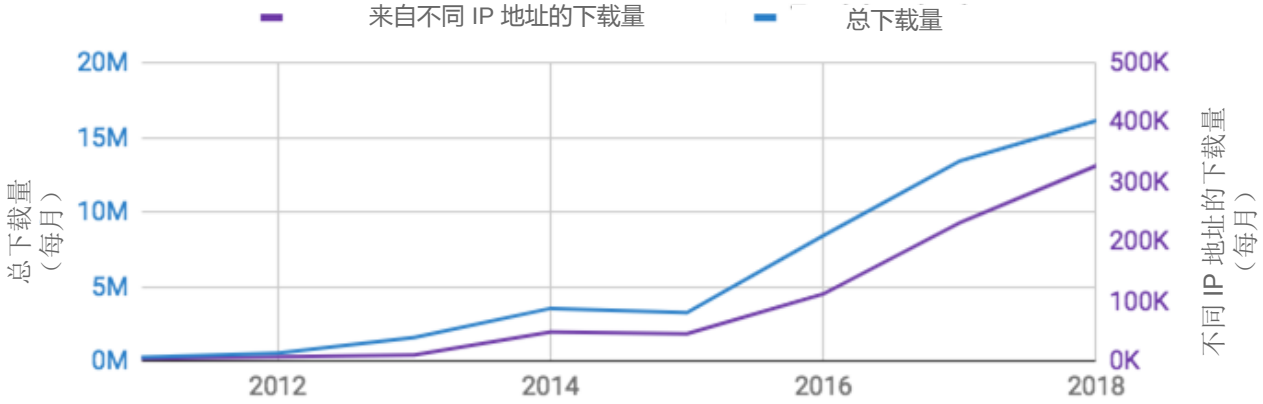
机器人软件下载情况

下图展示了 ROS.org 上机器人操作系统（ROS）二进制软件包的随时间变化的下载量。ROS 是一款使用广泛的开源机器人软件。很多商业制造商和学术研究者都在使用 ROS。左轴是每月的平均下载总数，而右轴是来自不同 IP 地址的平均月下载数。

自 2014 年以来，总下载数和不同 IP 地址的下载数分别增长了 352% 和 567%。这表明人们对机器人学的兴趣和机器人系统的使用量都在增长。因为不同 IP 地址的下载数的增速比总下载数快，所以我们可以推断 ROS 用户变多了，而不只是已有用户在更频繁地使用 ROS。

参见 附录 9 了解相关数据和方法。

ROS 软件包下载量（2011—2018）
来源：ROS 指标报告



自 2014 年以来，ROS.org 上不同 IP 地址的 ROS 软件包下载量增长了 567%

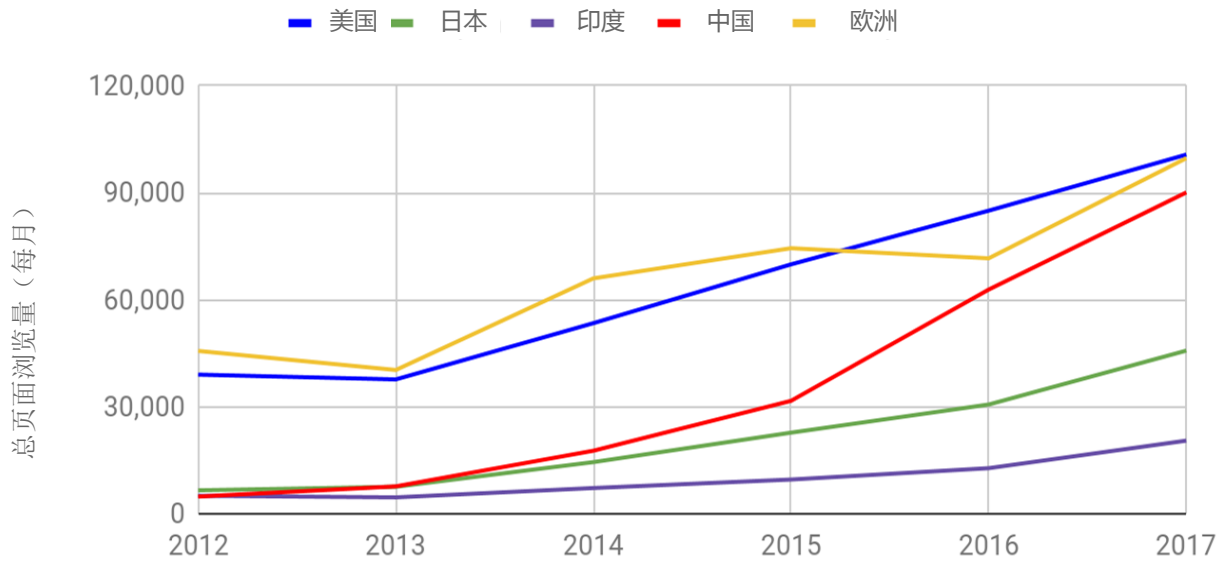
机器人软件下载情况

下图展示了自 2012 年以来浏览 ROS.org 页面次数最多的五个地区。

美国和欧洲的 ROS.org 页面浏览量最高。中国在经历了最大的增长之后也已经相差不多。中国 2017 年的浏览数量比 2012 年时多 18 倍。ROS.org 指出，中国浏览量的增长是自然增长，不是由于在中国的营销或其它资源的增长。

参见附录 9 了解相关数据和方法。

ROS.org 页面浏览量（2012-2017）
来源：ROS 指标报告



2017 年 ROS.org 在中国的页面浏览量比 2012 年时多 18 倍

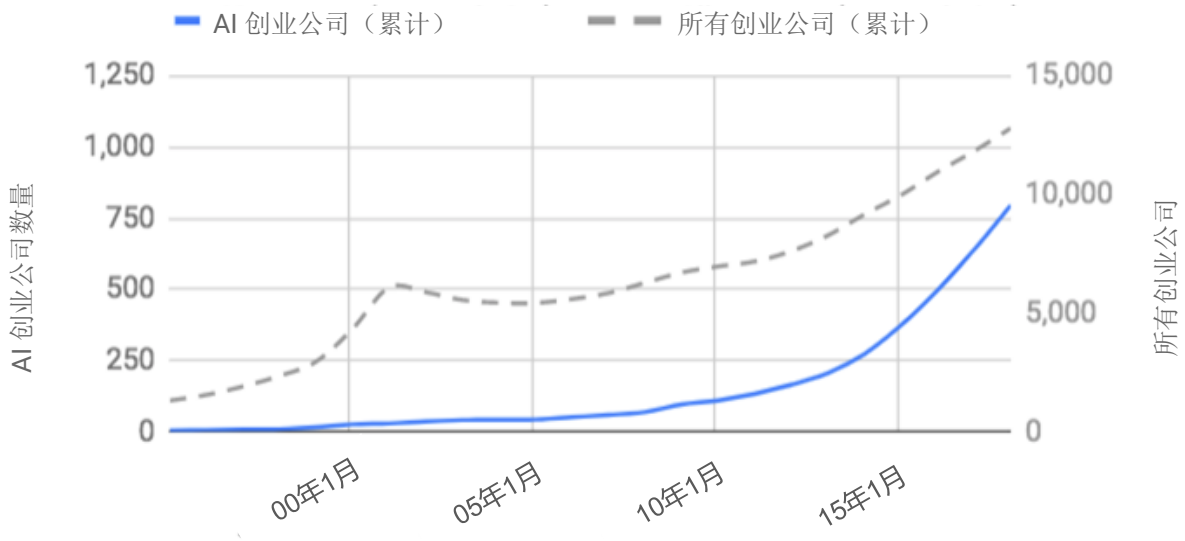
创业/投资：人工智能创业公司

下图展示了在各个年份活跃的获得了风险投资的美国私营创业公司的数量。蓝线（左轴）仅展示了 AI 创业公司，而灰线（右轴）则展示了所有获得了风险投资的创业公司——其中包括 AI 创业公司。这幅图给出了每年一月的创业公司总数。除了被移出数据集的创业公司实例（详情参见附录），创业公司的数量是逐年累计的。

从 2015 年 1 月到 2018 年 1 月，活跃的 AI 创业公司数量增加了 2.1 倍，而所有活跃的创业公司数量增加了 1.3 倍。大体而言，所有活跃创业公司的增长相对比较稳定，而 AI 创业公司的数量呈指数增长。

可参见 附录 10 了解相关数据和方法。

AI 创业公司（美国，1995 年 1 月 - 2018 年 1 月）
来源：Sand Hill Econometrics



注：上面展示的大部分 AI 创业公司都在开发 AI 系统。少量公司将 AI 用作自身业务的组成部分，但本身并不开发 AI 系统。参见附录了解更多细节。

从 2015 年 1 月到 2018 年 1 月，活跃的 AI 创业公司数量增加了 2.1 倍

同期，活跃的创业公司总数增加了 1.3 倍

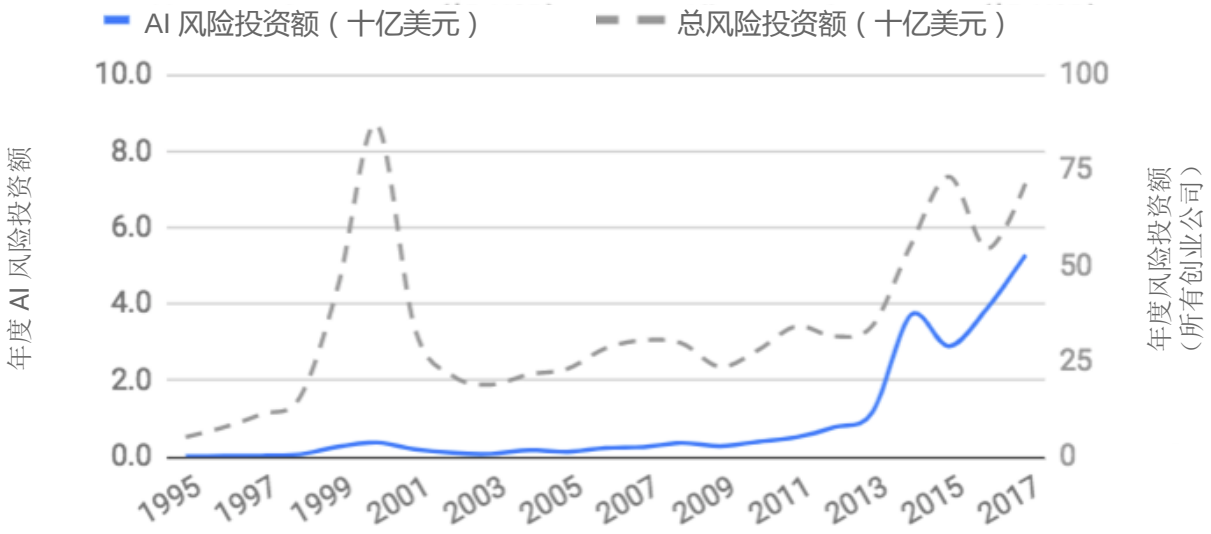
创业/投资：风险投资

下图展示了风险资本对活跃的美国创业公司的年度投资额，包含所有融资阶段。蓝线（左轴）仅展示了 AI 创业公司的融资情况，而灰线（右轴）则展示了所有获得了风险投资的创业公司的融资情况——其中包括 AI 创业公司。该数据是年度数据，并不是前一页那样的逐年累计的数据。

从 2013 年到 2017 年，AI 风险投资额增加了 4.5 倍，而所有风险投资额增加了 2.08 倍。1997-2000 年风险投资大爆发的原因是互联网泡沫。2014 和 2015 年更小规模的爆发则反映了一个经济增长相对较大的周期。

参见附录 10了解相关方法。

AI 创业公司的年度风险融资（美国，1995-2017）
来源：Sand Hill Econometrics



从 2013 年到 2017 年，AI 创业公司得到的风险投资额增加了 4.5 倍

同期，活跃的创业公司获得的风险投资总额增加了 2.08 倍

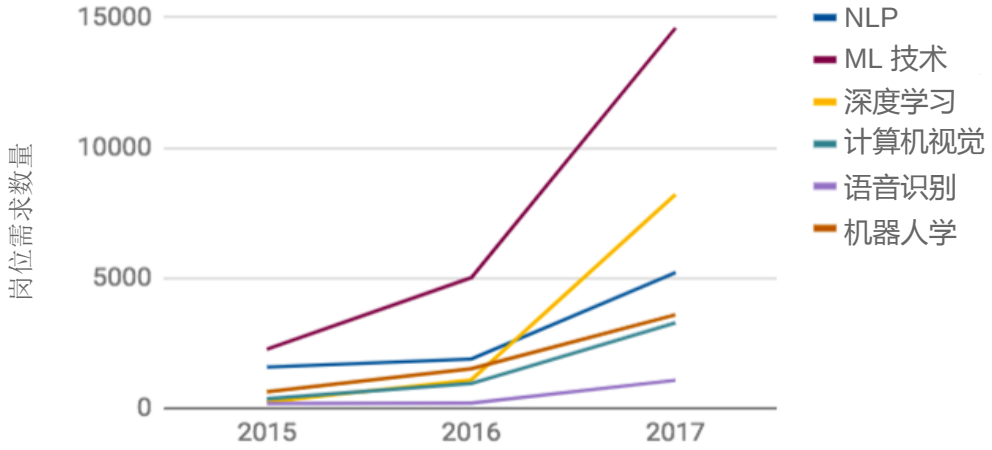
就业：需要 AI 技能的职位需求

下面的图片展示了每年需要 AI 技能的岗位需求的数量以及需要 AI 技能的岗位需求的增长情况。不同的 AI 技能并不互斥。

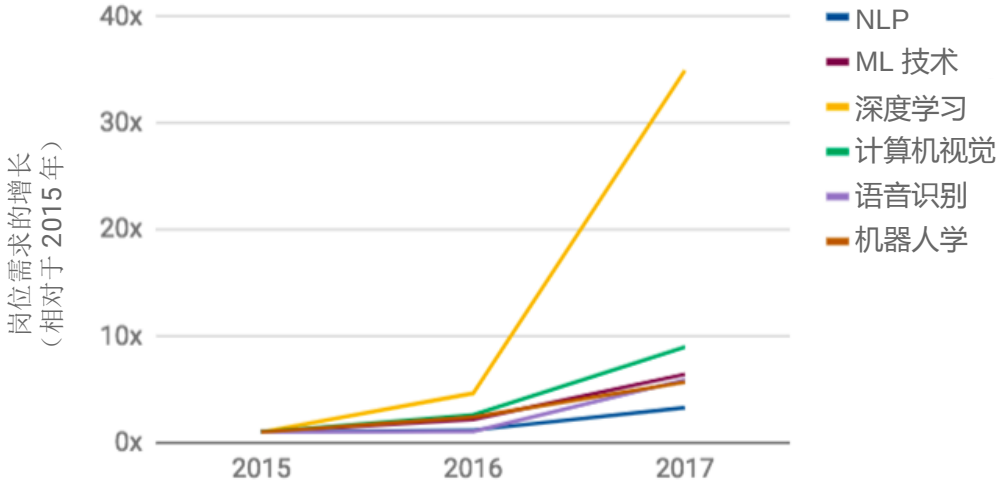
尽管机器学习（ML）是需求最多的技能，但深度学习（DL）技能需求的增长速度最快——从 2015 到 2017 年，需要深度学习技能的岗位需求数量增长了 35 倍。

参见 附录 11 了解相关方法。

根据所需 AI 技能统计的岗位需求（2015 — 2017）
来源：Monster.com



根据所需 AI 技能统计的岗位需求的增长（2015 — 2017）
来源：Monster.com



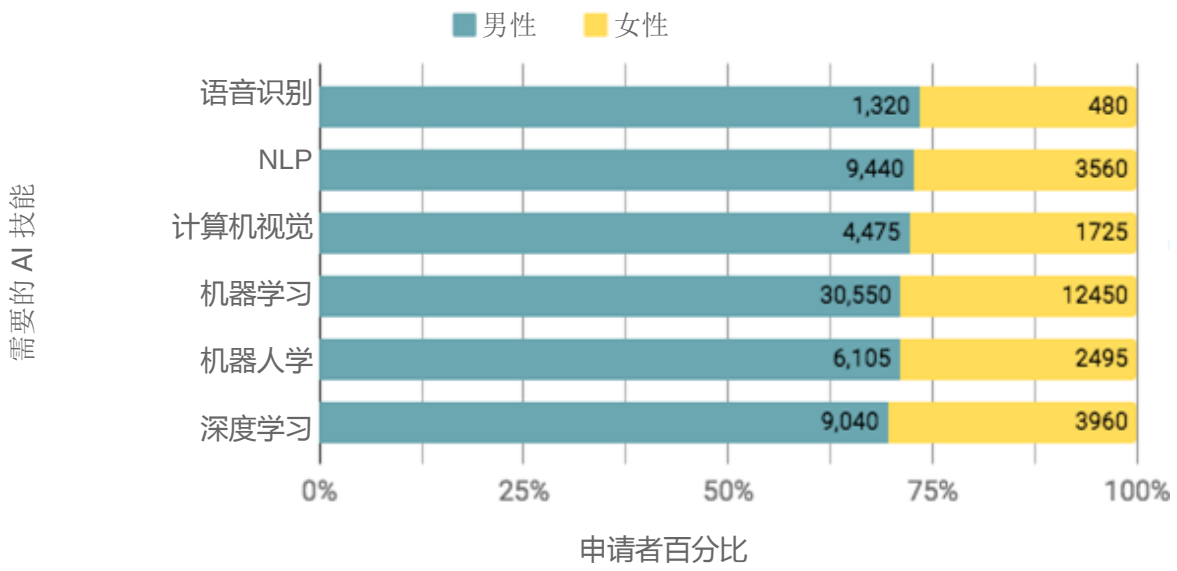
就业：工作申请者的性别多样性

下图展示了 2017 年 AI 工作岗位申请者的男性和女性数据。该数据是根据所需技能统计的，且技能之间不互斥。申请数量并不代表聘用数量，也不能代表该行业的整体情况。

平均而言，男性占美国 AI 工作申请者的 71%。因为机器学习需求的工作岗位有最多的申请者，所以该平均数受到了机器学习工作申请情况的很大影响。除了机器学习之外，深度学习和机器人学相比其它类别的性别多样性更高。

参见 附录 11 了解相关数据和方法。

按性别统计的工作申请者（2017）
来源：Gartner TalentNeuron



男性平均占美国 AI 工作申请者的 71%

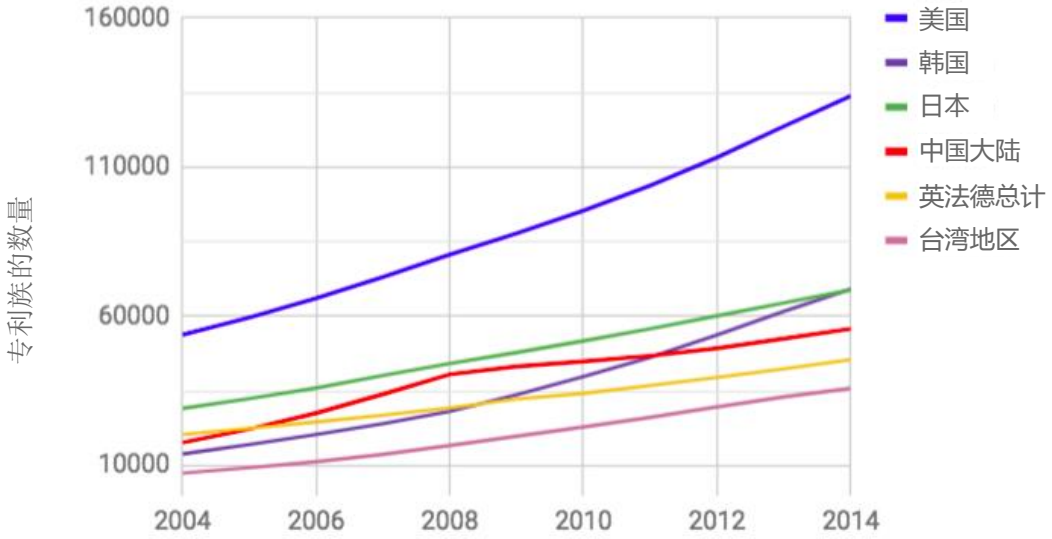
专利

下面的图片展示了按地区统计的 AI 专利的数量和增长情况。我们使用 IPC 代码对属于“认知和含义理解”以及“人机接口技术”领域的专利进行了汇总。跟踪专利随时间的变化情况是很困难的。与该指标相关的注释和细节说明请参阅附录。

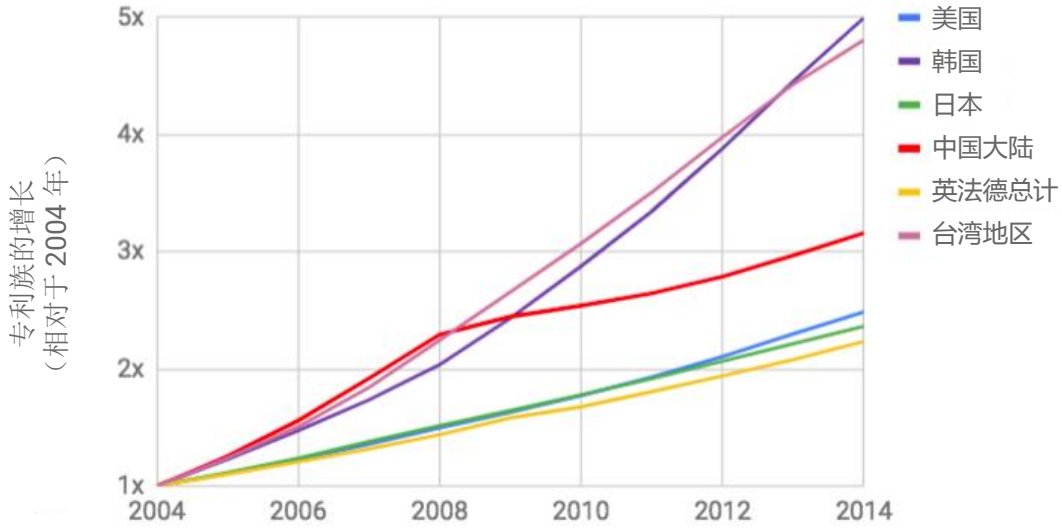
2014 年，大约 30% 的 AI 专利都源自美国，之后是韩国和日本，这两个国家分别持有 16% 的 AI 专利。在占比前列的地区中，韩国和台湾增长最显著，它们 2014 年的 AI 专利数量是 2004 年的近 5 倍。

参见 附录 12 了解相关数据和方法。

按发明者地区统计的 AI 专利（2004—2014）
来源：amplified



按发明者地区统计的 AI 专利的增长（2004—2014 ）
来源：amplified



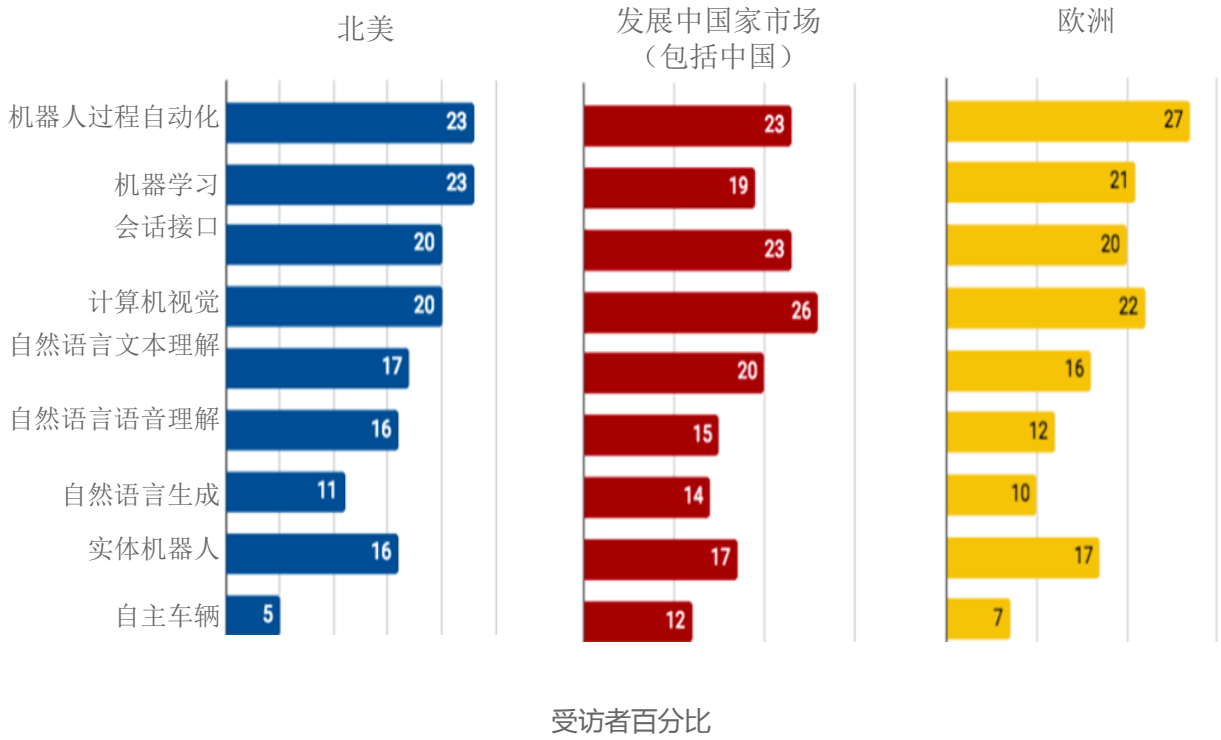
AI 采用情况：按地区统计

下面的图片和下一页的图片展示了 McKinsey & Company（麦肯锡）一项调查的结果，该调查有 2135 位受访者，每位受访者都代表他们所属的组织。图中给出了其组织至少在一个职能和业务部门应用了 AI 技术的受访者的百分比。受访者可以选择多个 AI 技术。下一页给出了亚太地区、印度、中东、北美和拉丁美洲的数据。

尽管某些地区比其它地区更多地采用了某些技术，但各个地区的 AI 技术应用相对均衡。我们希望未来能跟踪公司采用情况随时间的变化。

参见 附录 13 了解相关数据和方法。

至少有一项公司职能使用的技术（2018）
来源：McKinsey & Company



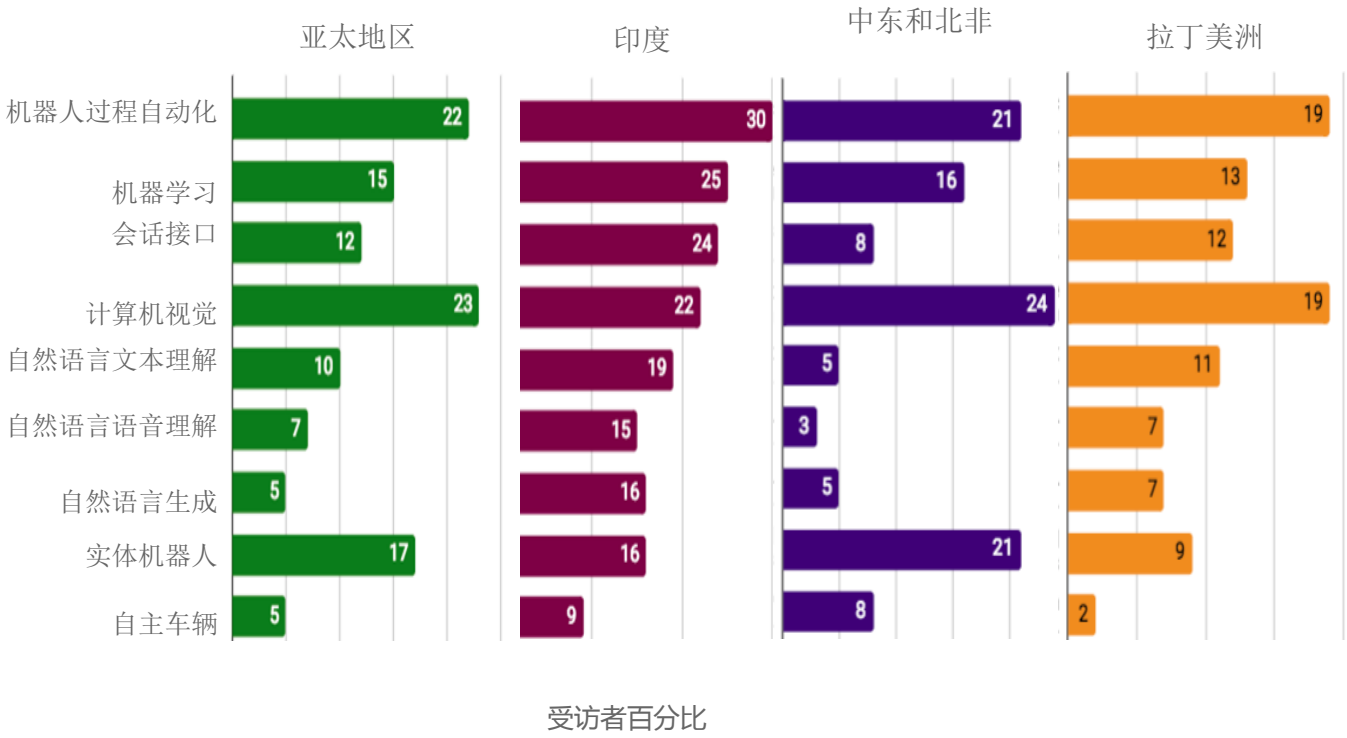
注： 北美：N = 479；发展中国家（包括中国）：N = 189 （*中国 N = 35*）；欧洲：N = 803

“……我们发现世界各地的不同行业、职能部门和地理位置都在广泛采用不同的AI 技术；大约有一半公司已经将 AI 纳入了一项企业业务过程。但是，现在仍处于早期阶段；大部分公司尚未采用大规模获取 AI 价值所必需的补充措施……”
——Michael Chui, McKinsey & Company

AI 采用情况：按地区统计（续）

对北美、发展中国家市场（包括中国）和欧洲的描述和数据见前一页。

至少有一项公司职能使用的技术（2018）
来源：McKinsey & Company



注： 亚太地区：N = 263；印度：N = 197；中东和北非：N = 77；拉丁美洲：N = 127

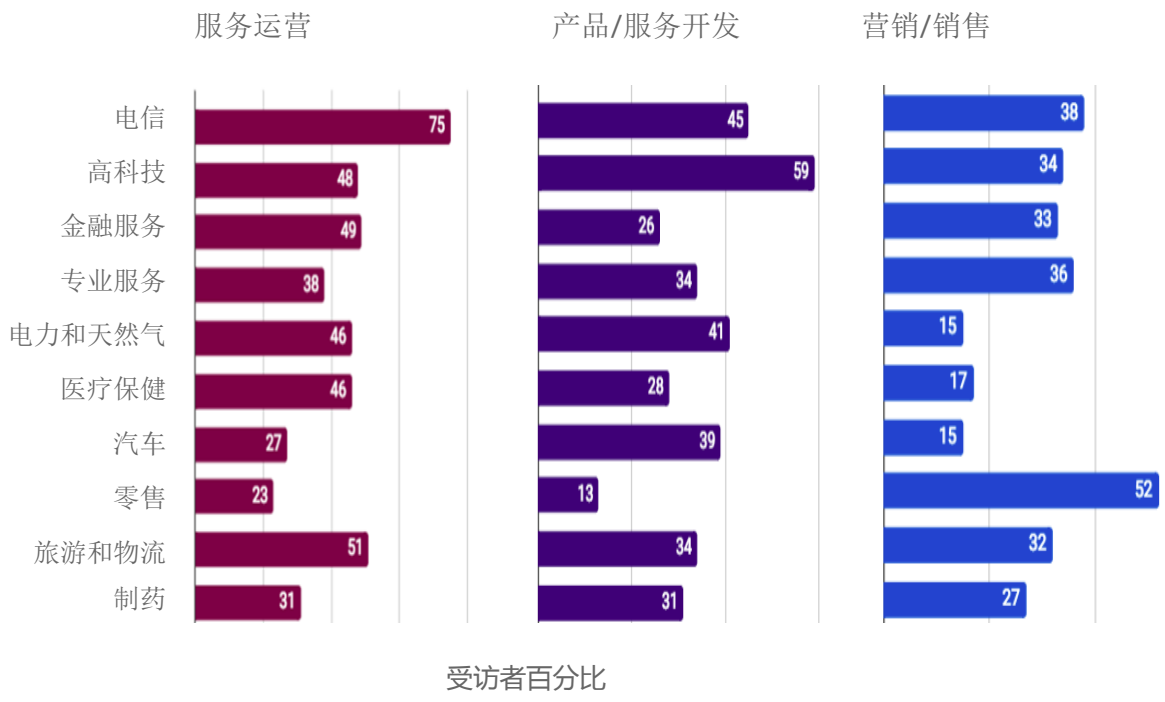
AI 采用情况：按行业和职能统计

下面的图片和下一页的图片展示了 McKinsey & Company（麦肯锡）一项调查的结果，该调查有 2135 位受访者，每位受访者都代表他们所属的组织。图中给出了其组织在某个业务职能中试点过或已纳入了 AI 能力的受访者的百分比。受访者可以选择多个 AI 能力。下一页给出了生产制造、供应链管理和风险管理的数据。

组织机构倾向于将 AI 能力整合进能在它们的行业中提供最大价值的职能中。比如，金融服务已经深度地将 AI 纳入了风险管理，而汽车行业则将其纳入了生产制造，零售业将其用在了营销/销售上。这说明针对特定应用（比如生产制造）的 AI 的发展速度很可能与该特定职能非常重要的行业的吸收能力息息相关。

参见 附录 13 了解相关数据和方法。

按行业和职能统计的 AI 采用情况（2018）
来源：McKinsey & Company

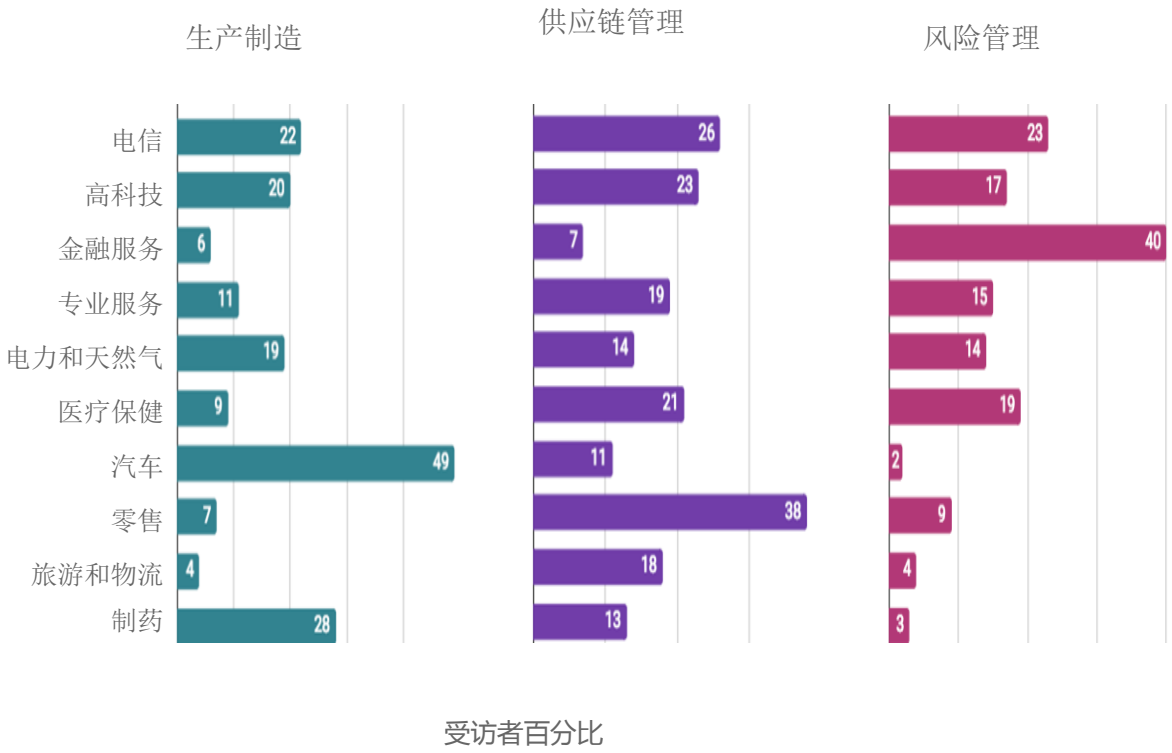


注：电信：N = 77；高科技：N = 215；金融服务：N = 306；专业服务：N = 221；电力和天然气：N = 54；医疗系统和服务：N = 67；汽车和组装：N = 120；零售：N = 46；旅游、运输和物流：N = 55；制药和医疗产品：N = 65。

AI 采用情况：按行业和职能统计（续）

服务运营、产品/服务开发、营销/销售的描述和数据见上一页。

按行业和职能统计的 AI 采用情况（2018）
来源：McKinsey & Company



注：电信：N = 77；高科技：N = 215；金融服务：N = 306；专业服务：N = 221；电力和天然气：N = 54；医疗系统和服务：N = 67；汽车和组装：N = 120；零售：N = 46；旅游、运输和物流：N = 55；制药和医疗产品：N = 65。

组织机构倾向于在能创造最大价值的业务职能中采用 AI

这说明针对特定应用的AI 的发展速度很可能与特定职能非常重要的行业的吸收能力息息相关

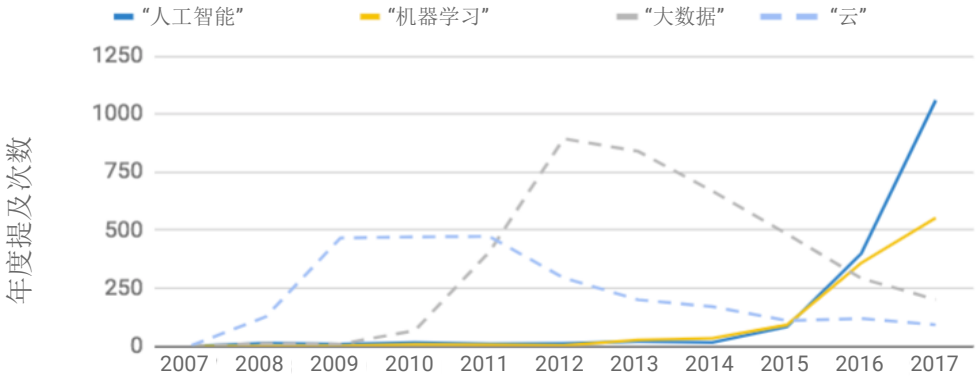
营收电话会议提及情况

下面的图片给出了按行业统计的在公司营收电话会议（*earnings calls*）中提到了“人工智能（AI）”和“机器学习（ML）”的情况。第一张图仅展示了 IT 行业公司的提及情况，因为该行业提及 AI 和 ML 的次数要多得多。第二张图展示了 IT 行业之外其它行业的提及情况。另外还添加了“大数据”和“云”的提及情况以供比较。该分析仅使用了在纽约证券交易所上市交易的公司。

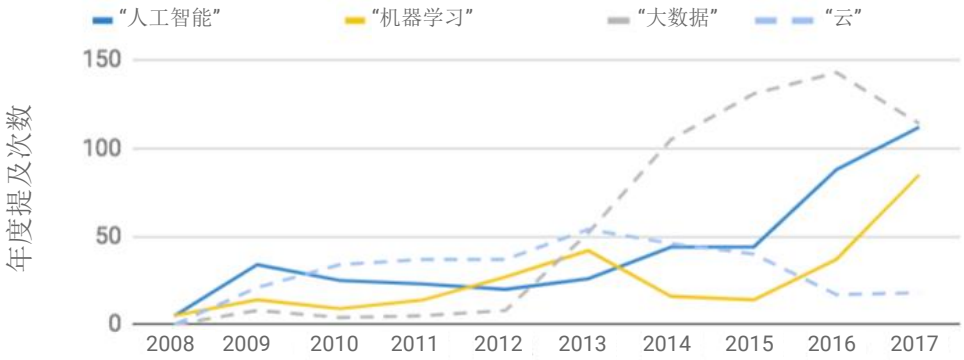
2015 年，IT 行业提及 AI 和 ML 的次数显著增长。对于其它大部分行业，增长是从 2016 年开始的。IT、非必需消费品、金融和医疗保健行业在营收电话会议中提到 AI 的次数最多。

参见 附录 14 了解相关数据和方法。

公司营收电话会议提及情况——IT 公司（2007-2017）
来源：Prattle



公司营收电话会议提及情况——其它公司之和（2008-2017）
来源：Prattle



提及 AI 次数最多的 10 家公司（2017）
英伟达 (93), LivePerson (50), Pegasystems (44), Facebook (40), Axon Enterprise (37), salesforce.com (36), 英特尔 (35), 微软 (34), Genpact Limited (34), Applied Materials (33)
提及 ML 次数最多的 10 家公司（2017）
Alphabet(57), Nasdaq (26), ServiceNow (21), Progress Software Corporation (21), Cadence Design Systems, Inc. (20), Splunk Inc. (18), Twitter (17), Overstock.com (14), Synopsys, Inc. (14), Aspen Technology (14)

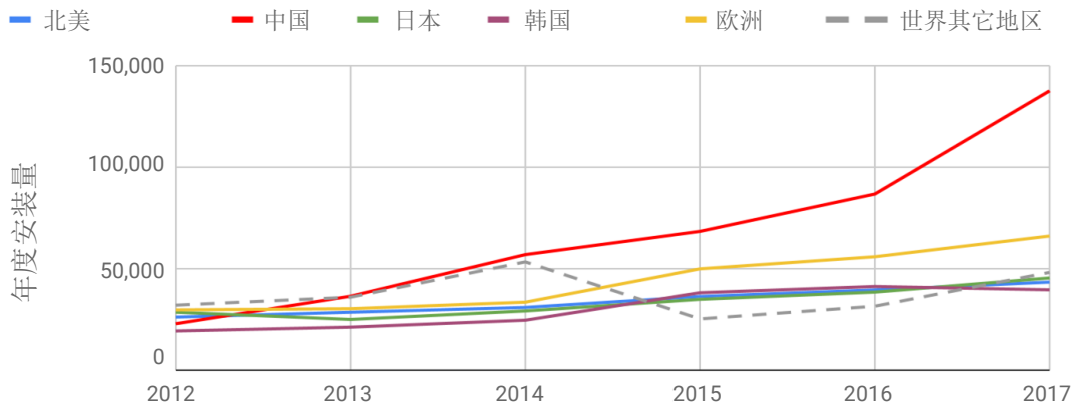
机器人安装情况

下面的图片展示了不同地区各年度的工业机器人安装情况。第一张图展示了机器人安装最多的五个地区，第二张图呈现了世界上其它地方的情况。

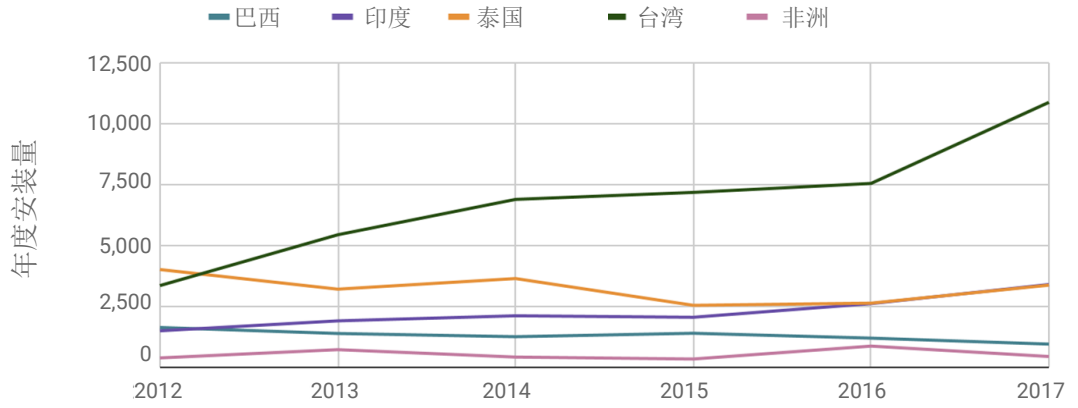
自 2012 年以来，中国的年度机器人安装数增加了 500%，而韩国和欧洲的增长分别为 105% 和 122%。在安装数量更少的地区，台湾地区的年度安装量最高，而且 2012-2017 年的增长也最大。

参见 附录 15 了解相关数据和方法。

安装机器人最多的五个地区（2012-2017）
来源：ifr.org



世界其它地区的安装情况（2012-2017）
来源：ifr.org



注：安装数量是指国内生产数加上进口数再减去出口数。

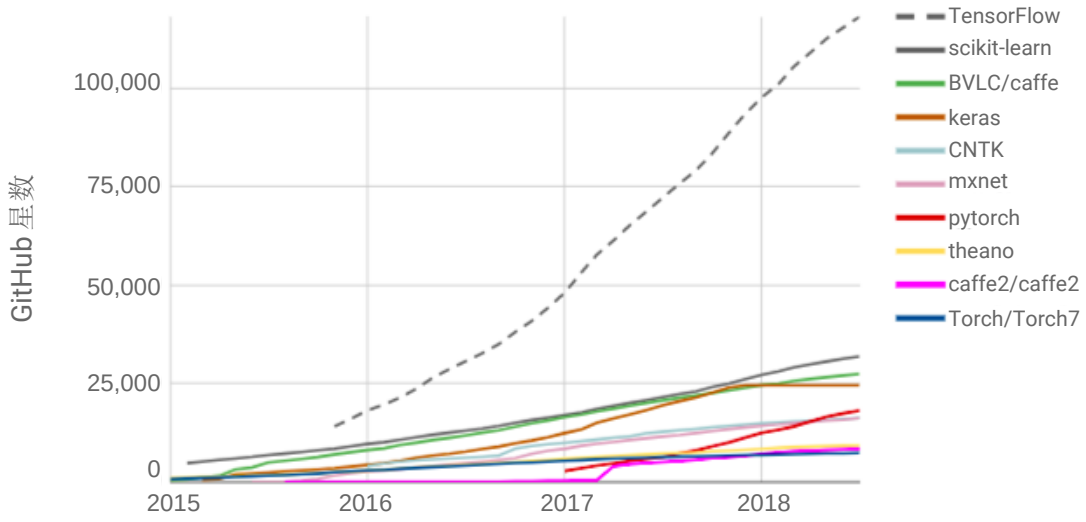
GitHub 星数

下面的图片展示了各种 AI 和 ML 软件包在 GitHub 上被标星的次数。这能让我们粗略地估计各种 AI 编程框架的流行程度。

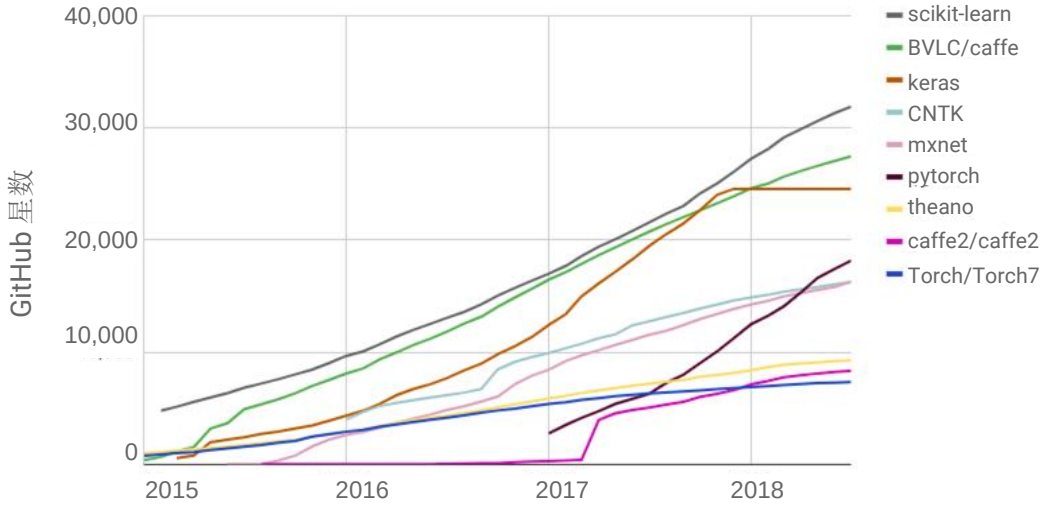
近期出现了两个趋势，一是主要公司支持的框架的流行程度越来越高（即 TensorFlow（谷歌）、Pytorch（Facebook）、 mxnet（亚马逊）），二是 TensorFlow 相对其它语言实现了显著的普及。

参见 附录 16 了解相关数据和方法。

各 AI 软件库的累计星数（2015-2018）
来源：GitHub



除 TensorFlow 之外的各 AI 软件库的累计星数（2015-2018）
来源：GitHub



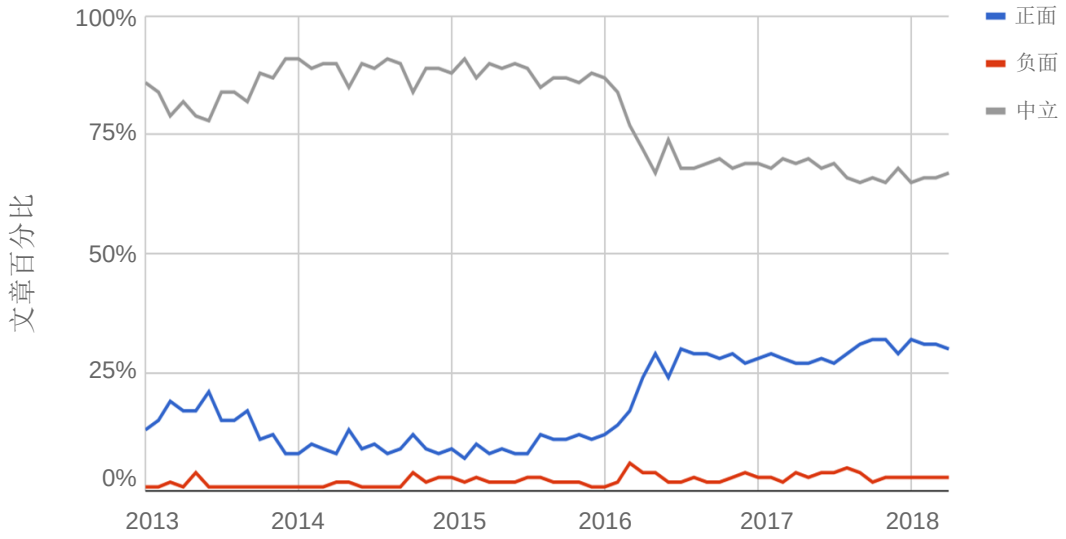
媒体报道对 AI 的情绪

下图展示了包含“人工智能”术语的大众媒体文章被分类为正面、负面或中立的百分比。

提到 AI 的文章的情绪中立性降低了，正面性提高了，尤其是自 2016 年初以来——正面文章比例从 2016 年 1 月的 12% 提升到了 2016 年 7 月的 30%。自那以后，正面文章的占比一直保持在 30% 附近。

参见 附录 17 了解相关数据和方法。

谈到了 AI 的文章的情绪（2013—2018）
来源：TrendKite



2016 年到 2018 年，正面的 AI 相关文章比例增大了 2.5 倍

政府提及情况

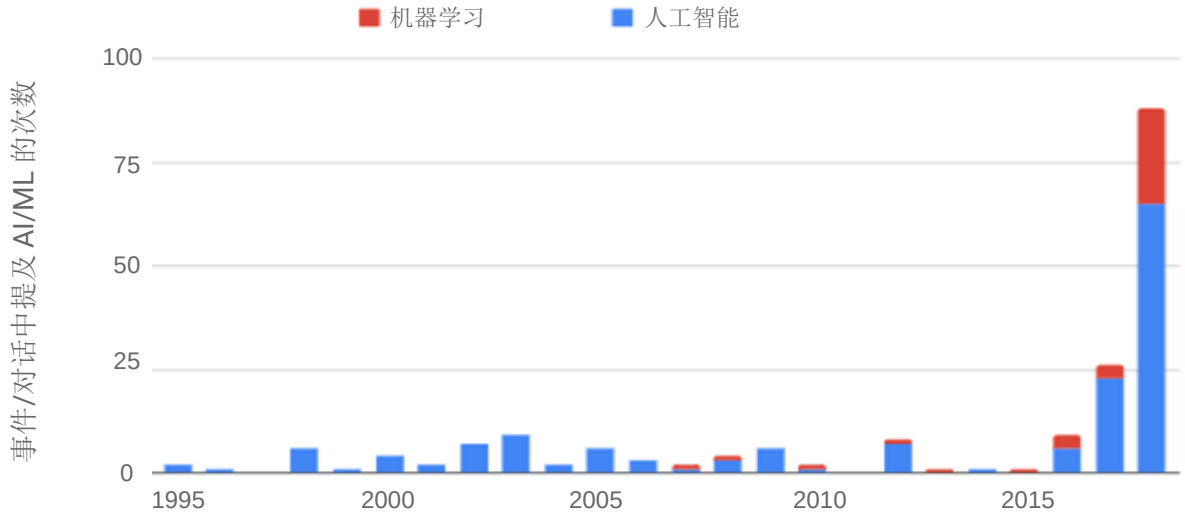
下图以及下一页的图片展示了美国国会记录以及加拿大议会和英国议会的议程记录（也被称为 Hansards）的转录文本中提到了术语“人工智能”和“机器学习”的次数。

自 2016 年以来，这三个政府提到这些术语的次数大幅增长。此外，在这三个国家，“机器学习”在 2016 年之前都鲜有提及，相对于“人工智能”，其在总提及次数中的比重也较小。

注意，统计口径的差异让我们难以比较各个国家。比如，美国数据中的每次计数都表示“机器学习”或“人工智能”在一个事件或对话中被至少提及了一次。而英国数据的每次计数则表示“机器学习”或“人工智能”至少在一个评论中被提及了一次。所以，我们没有在国家之间进行比较，而是给出了各个国家随时间变化的趋势。

以下是美国的数据，加拿大和英国的数据见下一页。参见 附录 18 了解相关数据和方法。

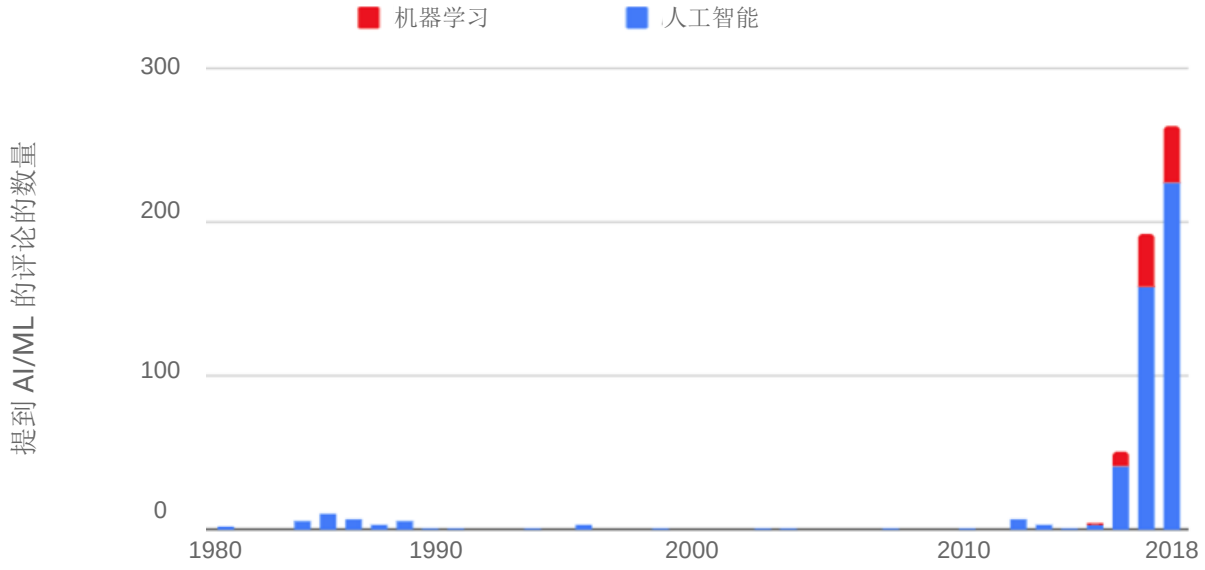
美国国会提到 AI 和 ML 的次数（1995-2018）
来源：美国国会记录网站、麦肯锡全球研究院的分析



注：本分析包含了美国参议院和众议院的数据。数据截止至 2018 年 11 月 20 日。

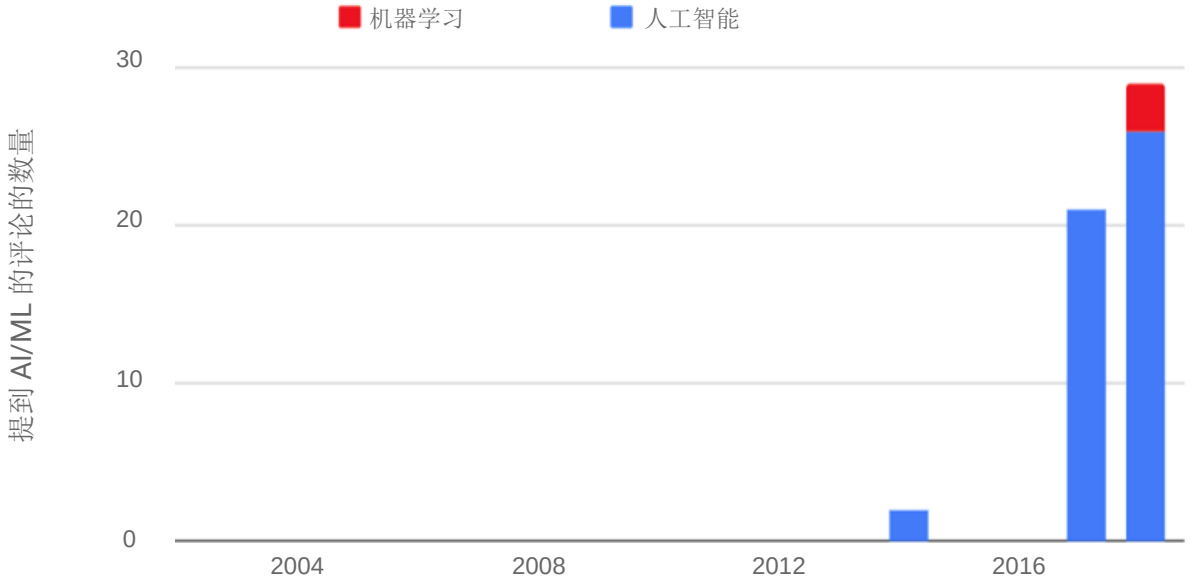
政府提及情况

英国议会提到 AI 和 ML 的次数（1980-2018）
来源：英国议会网站、麦肯锡全球研究院的分析



注：这个分析包含了英国下议院、上议院、威斯敏斯特大厅和委员会的数据。2018 年的数据截止至 2018 年 11 月 20 日。

加拿大议会提到 AI 和 ML 的次数（2002-2018）
来源：加拿大议会网站、麦肯锡全球研究院的分析



注：本分析仅包含加拿大下议院的数据。2018 年的数据截止至 2018 年 11 月 20 日。请注意这是非商业用途的非正式重现。想要使用这张表的读者应该注意议长的许可并不涵盖获利性质的商业用途。

技术性能

目标检测：ImageNet

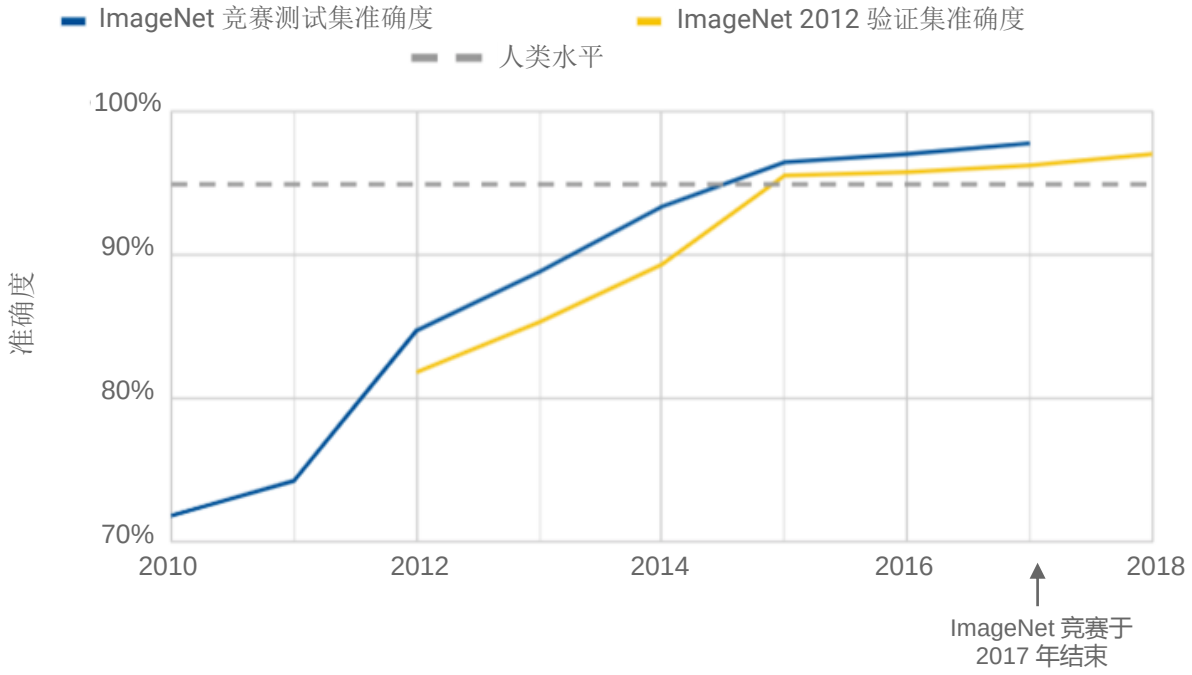
下图展示了 ImageNet 准确度随时间的变化情况。ImageNet 竞赛一直持续到了 2017 年，分数是在一个留存的专用于竞赛的“测试”数据集上得到的。因为竞赛已经不再举办，所以人工智能指数研究选择通过研究论文跟踪 ImageNet 方面的持续进展，这是在 ImageNet 2012 语料库的验证数据集上测试的。

因为我们是根据验证数据集绘制了结果且没有了竞赛，所以不再有可以用以评估模型的留存测试数据集。为了带来发展进步的感觉，我们给出了 2012 年到 2018 年在 ImageNet 2012 语料库上的突出结果。如图所示，ImageNet 表现在持续提升。

这一指标还凸显了建模随时间的 AI 进展的一大固有难题：如果可以围绕一个竞赛构建一个特定的研究指标，那么该竞赛的退休将使得人们难以了解真正的进展情况。但是，由于其数据集是开放可用的，所以能在一定程度上保证连续性。

参见 附录 19 了解相关数据和方法。

ImageNet（2010 –2018）
来源：ImageNet；参见附录



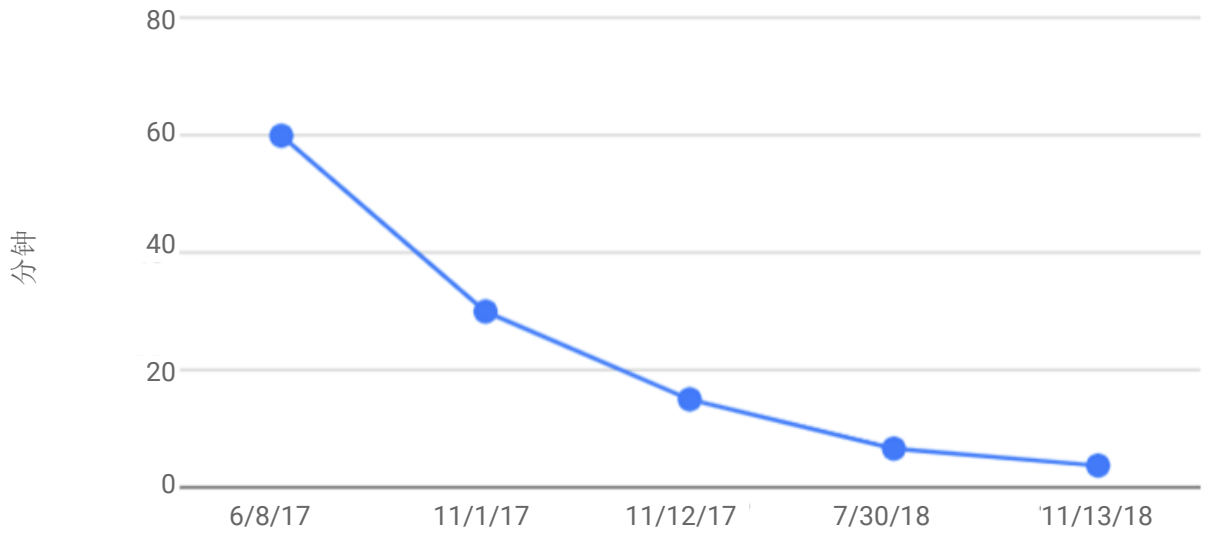
目标检测：ImageNet 训练时间

下图展示了训练一个能以高准确度分类 ImageNet 语料库（一个图像数据库）的图像的网络所需的时间。该指标能够代表 AI 领域那些资源充足的参与者训练大型网络执行 AI 任务（比如图像分类）所需的时间。因为图像分类是一个相对通用的监督学习任务，因此这一指标的进展也可对应其它 AI 应用的更快的训练时间。在一年半的时间里，训练一个网络所需的时间已经从大约 1 小时降至了大约 4 分钟。

ImageNet 训练时间指标还能反应 AI 研究的产业化程度。能够降低 ImageNet 训练时间的因素包括：算法创新和基础设施投资（比如，用于训练系统的底层硬件或用于将硬件连接到一起的软件）。

参见 附录 19 了解相关数据和方法。

ImageNet 训练时间（2017 年 6 月- 2018 年 11 月）
来源：arXiv.org；参见附录了解作者信息



2018 年 11 月的 ImageNet 训练时间比 2017 年 6 月时快 16 倍

实例目标分割：COCO

下图展示了在背景中的常见目标挑战赛（COCO）的表现。

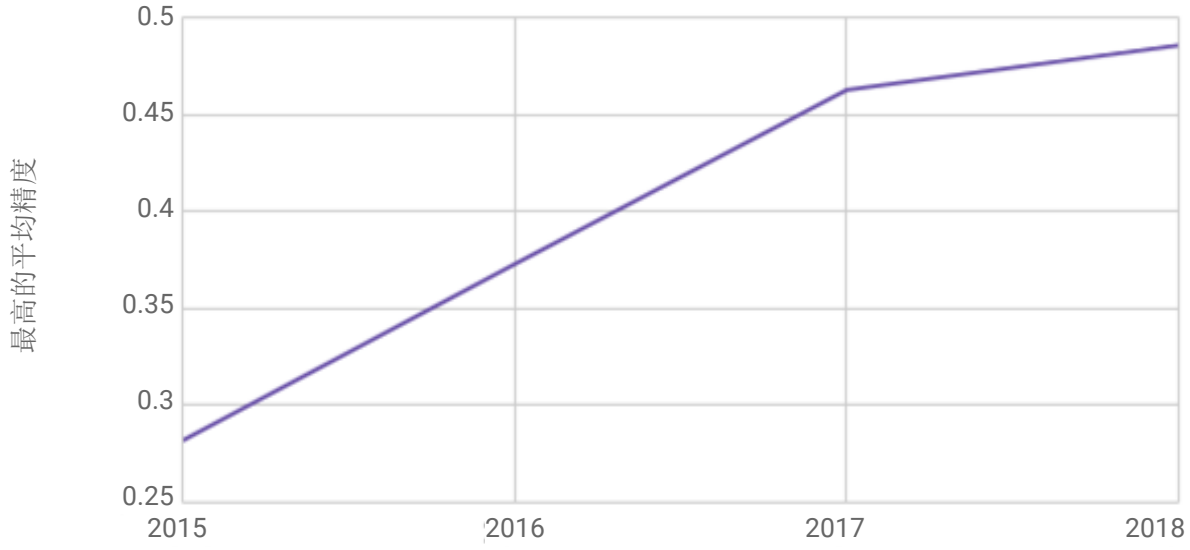
2017 年，在计算机视觉算法在 ImageNet 提供的目标检测和图像分类任务上取得了极好的表现之后，ImageNet 挑战赛退出了舞台。自那以后，研究社区转向了更难的计算机视觉任务。该社区已经将研究重心转向了需要更复杂推理的视觉任务，比如以像素级准确度定位目标位置（被称为目标实例分割）以及以像素级精度将一个场景分成不同的区域（被称为形义分割）。

COCO 指标能体现上面提到的细节性计算机视觉任务的表现。该挑战赛的目标是构建能够定位每个相关目标以及准确划定目标边界的算法。

自 2015 年以来，在 COCO 挑战赛上的最高平均精度已经提升了 0.2 点，即提升了 72%。

基础数据参见 这里，或了解有关 COCO 挑战赛的更多信息。

COCO 实例分割挑战赛（2015-2018）
来源：COCO 排行榜



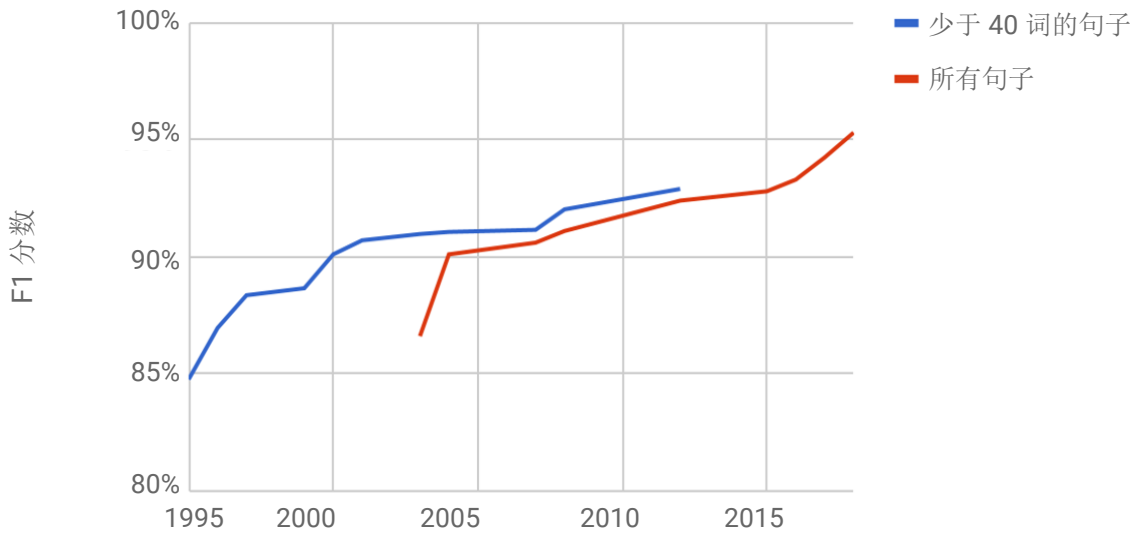
句法分析

下图展示了 AI 系统在确定句子的句法结构的任务上的表现。

句法分析指标是在特定任务（比如问答任务）中理解自然语言的第一步。最初的做法是使用类似于用于解析编程语言的算法，现在则基本都使用深度学习来做。自 2003 年以来，分析所有句子的 F1 分数已经增长了 9 个百分点（即增长了 10%）。

参见 附录 20 了解相关数据和方法。

成分句法分析——Penn Treebank（1995–2018）
来源：参见附录



从 2013 年到 2018 年，成分句法分析的表现水平提升了 10%

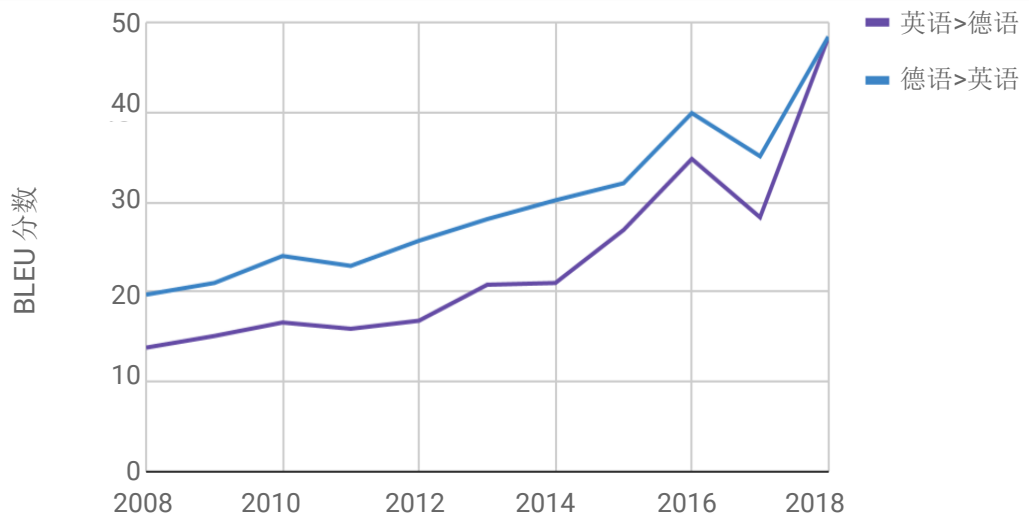
机器翻译

下图展示了 AI 系统在英语到德语以及德语到英语的新闻翻译任务上的表现。

现在的英德翻译 BLEU 分数比 2008 年时高 3.5 倍。同期德英翻译高 2.5 倍。因为每一年都会使用不同的测试集，所以各年的分数比较并不是完善的（我们相信这也是 2017 年分数下降的原因——更多详情见附录）。尽管如此，BLEU 分数仍能代表机器翻译的进展。

参见 附录 21 了解相关数据和方法。

新闻翻译——WMT 竞赛（2008-2018）
来源：EuroMatrix



英语翻德语的 BLEU 分数比 2008 年时高 3.5 倍

"……表现领先的系统在一些方面做到很好。它们利用了用于神经机器翻译的 transformer 架构，并且还能更高效地利用数据。transformer 架构能比更早期的神经机器翻译模型（基于 RNN）得到更好的结果，但需要仔细谨慎的超参数选择和优化……"
—Barry Haddow，爱丁堡大学

商用机器翻译的性能表现
更多有关商用机器翻译发展情况的信息请参阅这里。

问答：ARC

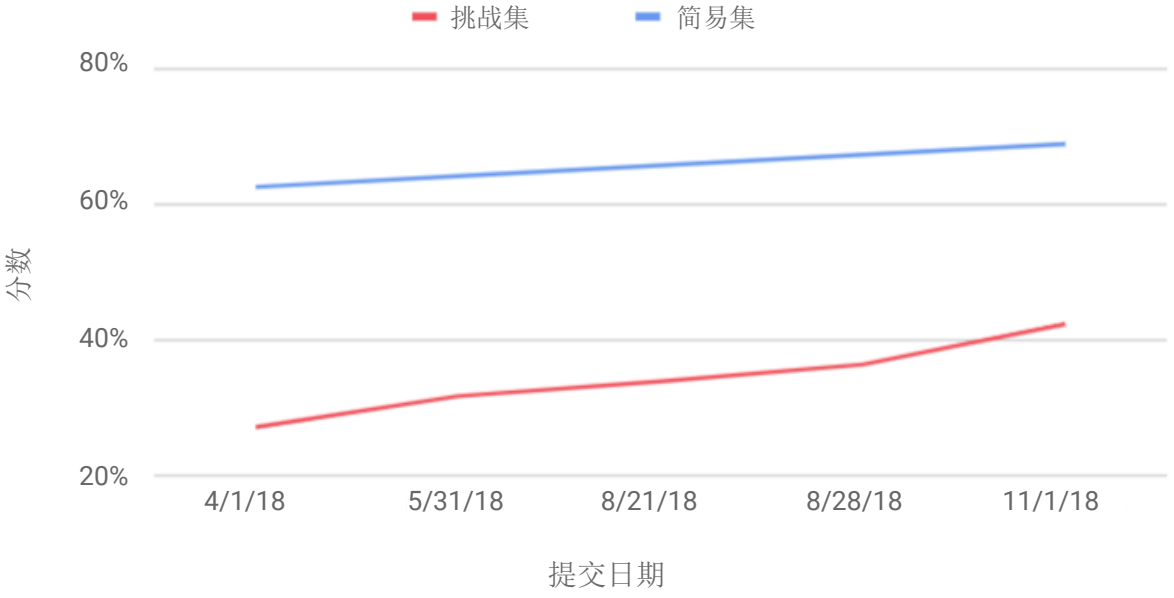
下图展示了 AI 系统在 AI2 推理挑战赛（ARC）上的表现。ARC 数据集包含 7787 个真正的小学水平（美国三到九年级）的多项选择科学问题，该挑战赛的目的是鼓励人们研究高级问答任务。根据问题难度，该数据集分为挑战集（2590 个问题）和简易集（5197 个问题）。

挑战集仅包含基于检索的算法和词共现算法都回答错误的问题。这些问题都是文本形式的英语的考试问题，包含其文件中指出的多个年级水平。每个问题都有多项选择的结构（通常为 4 个答案选项）。伴随这些问题的是 ARC 语料库，这是一个包含 1400 万无序的科学相关句子的聚合。不能确保这些问题的答案能够在该语料库中找到。

ARC 基准于 2018 年 4 月发布。在 2018 年，AI 系统在简易集上的表现已经从 63% 提升到了 69%，在挑战集上的表现已经从 27% 提升到了 42%。

参见 附录 22 了解相关数据和方法。

ARC 排行榜（2018 年 4 月 – 2018 年 11 月）
来源：艾伦人工智能研究所



注：这幅图展示了领先的提交结果，并用趋势线进行了连接。更多有关提交日期的信息请参阅附录。

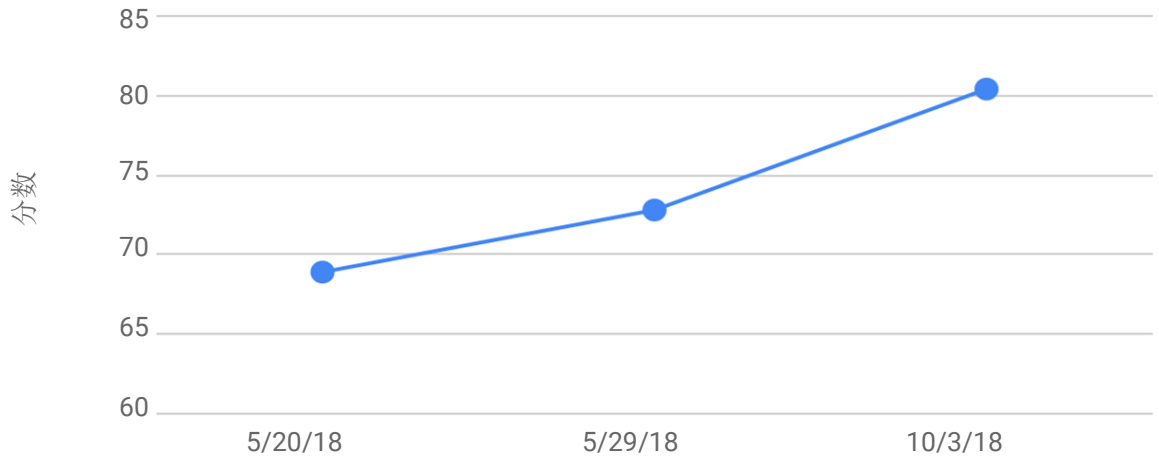
问答：GLUE

下图展示了来自 GLUE 基准排行榜的结果。通用语言理解评估（GLUE）是一个意在测试自然语言理解（NLU）系统的新基准，该基准涉及多种任务，旨在鼓励开发不局限于特定任务的系统。GLUE 基准由 9 个子任务构成——两个针对单句的任务（衡量语言可接受性和情感）、三个针对相似性和释义的任务、四个针对自然语言推理的任务，包括 Winograd 模式挑战赛。不同任务的语料库规模从少于 1000 到超过 400 000 不等。度量指标包括准确度/F1 和马修斯相关系数。

尽管该基准 2018 年 5 月才发布，但自发表基线成果以来，AI 系统的表现与估计的 90% 的非专家人类水平的差距已经缩减了大约一半。

参见 附录 23 了解相关数据和方法。

GLUE 基准排行榜（2018 年 5 月 – 2018 年 10 月）
来源：Gluebenchmark.com



"……我们尚未看到围绕 GLUE 的大型社区出现，但我认为我们已经有两个重大的成果提交可被视为句子理解的表征学习的重大里程碑。当前最佳的谷歌 BERT 尽管刚出一个多月，就已经被引用了 8 次。这是近期的 NLP 会议 EMNLP 上走廊讨论的主要焦点，看起来它将成为近期未来语言理解问题研究所要击败的标准基线，无论该研究是否与预训练的表征有关……"
—Sam Bowman，纽约大学

其它问答数据集/任务
问答数据集的数量正在快速增长。人工智能指数2018年度报告未包含一些其它数据集：Squad 2.0; OpenBookQA; HotpotQA; MultiRC; NarrativeQA; DrQA; CoQA; QuAC; CWQ (rc setting)

其它指标

衍生指标

衍生指标部分的目标是呈现前文中所展现出的趋势之间的关系。

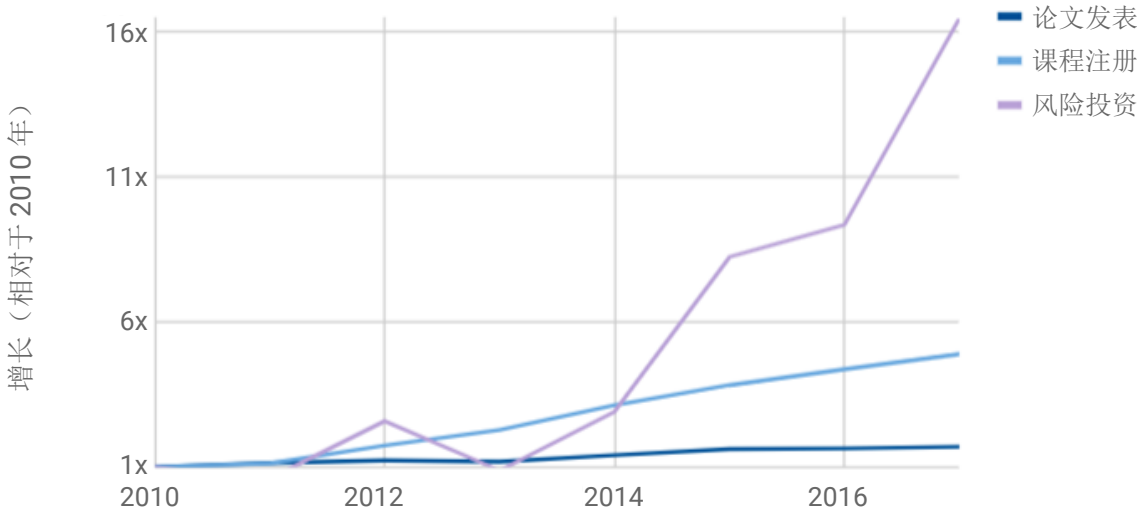
第一个指标是学术界-产业界动态，给出了选定的学术界指标与选定的产业界动态之间的增长情况对照。第二个指标是人工智能活力指数，即将学术界和产业界指标结合成的单个指数。

学术界-产业界动态

为了探索学术界和产业界中 AI 相关活动之间的关系，我们首先从之前章节中选择了几个代表性指标。具体而言，我们检查了 Scopus 上的 AI 论文发表情况、几所美国大学的入门级 AI 和 ML 课程综合注册情况、对 AI 相关创业公司的风险投资情况。

这些指标的度量是无法直接进行比较的。为了分析趋势之间的关系，我们以 2010 年的数据为基准，对每个指标进行了标准化，展示了各指标的增长情况而非绝对数值。

学术界-产业界动态（2010 — 2017）
来源：Sand Hill Econometrics、Scopus、大学提供的数据



注：2017 年的报告仅包含斯坦福大学的课程注册指标，今年我们包含这个列表中的美国大学。

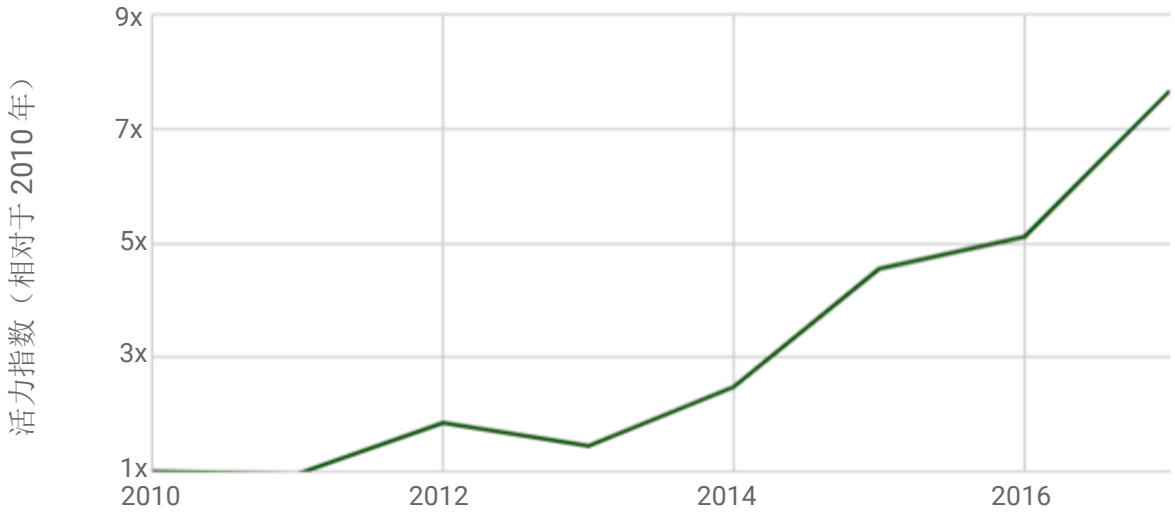
人工智能活力指数

AI 活力指数将三个学术界和产业界指标（论文发表、课程注册和风险投资）聚合成了单个度量。我们使用 AI 活力指数来量化 AI 领域的活力。类似于学术界-产业界动态，AI 活力指数基于 2010 年的数据进行了标准化。

AI 活力指数很大程度上由风险投资推动，因为相对其它两个指标（见上一页），风险投资增长最显著。这三个指标是同等加权的。我们鼓励读者下载我们的数据并调整这些指标和权重，以创造最符合自己对这些指标的相关性的特定看法的活力指数。

论文发表：见发表情况附录
课程注册：见注册情况附录
风险投资：见风险投资附录

AI 活力指数（2010 — 2017）
来源：Sand Hill Econometrics、Scopus、大学提供的数据



近期政府计划

下面列出的资源来自欧盟委员会的联合研究中心和 Tim Dutton 2018 年 6 月的《国家人工智能战略概述》

美国	欧洲 *（仅包含欧盟、英国和法国，未包含其它欧洲国家的一些计划）*	中国
2016 年 10 月：奥巴马政府提出了一项国家人工智能研发战略，目标是为研究投资，开发人与 AI 的合作方法，解决 AI 的安全、伦理道德、法律和社会影响问题，为 AI 训练创建公共数据集，通过标准和基准来评估 AI 技术。 **2018 年 5 月**：特朗普政府的 Summit on AI 宣布了其目标：（1）维持美国在 AI 领域的领导地位；（2）支持美国的工作者；（3）促进公共研究；（4）消除创新的阻碍。创建了人工智能专责委员会以为白宫提供有关 AI 研发优先级的建议以及考虑建立与学术界和工业界的合作关系。	**欧盟** **2018 年 4 月**：欧盟成员国签署了《人工智能合作宣言（Declaration of Cooperation on AI）》，同意共同合作。欧盟委员会发布了《人工智能通报（Communication on AI）》，目标是（1）促进欧盟的研究投资；（2）为社会经济变革做准备；（3）构建一个符合道德伦理与法律的框架。该通报称将投入 15 亿欧元（约 17 亿美元）来支持 2018-2020 年期间的 AI 研究。这是在 Horizon 2020 计划的 26 亿欧元基础上新增的投资。欧盟委员会的目标是在未来十年内向 AI 研究投入 200 亿欧元。 **2018 年 6 月**：欧盟委员会提出了数字欧洲（Digital Europe）计划，为 2021-2027 年指派了 92 亿欧元的预算（约 104 亿美元）。该计划侧重于推进 AI 技术以及确保 AI 在社会和经济中的广泛应用。欧盟委员会还提出开发共同的“欧洲算法库”，让所有人都能访问。	自 2014 年以来，中国政府已经推出了一系列重点国家人工智能计划，目标是到 2018 年在中国创造一个 147 亿美元的 AI 市场以及确保中国在 2030 年之前在 AI 领域领先世界。 **2015 年 7 月**：提出关注中国智能制造的互联网+计划。AI 方面的具体目标包括：提升公众对 AI 开发的支持；推动 AI 在智能家居、智能汽车和机器人等领域的普及；构建一个大型训练数据库，包括语音、图像、视频、地图和科研；开发和产业化计算机视觉、语言处理和人机交互等关键的 AI 技术。 **2016-2020**：《机器人产业发展规划》为中国的智能工业机器人和服务机器人的发展指出了方向。

注：更多信息见下一页。

近期政府计划（续）

下面列出的资源来自欧盟委员会的联合研究中心和 Tim Dutton 2018 年 6 月的《国家人工智能战略概述》

美国	欧洲 *（仅包含欧盟、英国和法国，未包含其它欧洲国家的一些计划）*	中国
2018 年 9 月：美国国防高级研究计划局（DARPA）宣布了一项 20 多亿美元的投资计划，以克服在 AI 技术上的局限性。AI Next 计划开始。美国众议院监督和政府改革委员会信息技术小组委员会发表了一份关于人工智能及其对政策的影响的白皮书《机器的崛起：人工智能及其对美国政策日益增长的影响》。 2017 年，美国在 AI 开发/投资方面的核心角色是私营公司，亚马逊和 Alphabet 这两家私营科技公司在研发方面各自投入了 161 亿美元和 139 亿美元。对比来看，美国国家科学基金会（NSF）加上 DARPA 与美国交通运输部（DOT）一起在自动化和无人系统方面的 2019 年总预算为 53 亿美元。	**法国** **2018 年**：法国近期发布了三份报告：（1）法国人工智能计划（French AI Plan）提出了一项有关研究、教育和创新的战略；（2）一份来自科学与技术议会办公室的报告，关注的是社会和监管问题；（3）Mission Villani 报告关注的是政策和减少人才流失，并强调了人工智能研究中多样性的重要性。 **英国** **2016 年**：英国议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。 **2018 年 4 月**：政府发布了《人工智能行业新政（AI Sector Deal）》，将投资 9.5 亿英镑（约 12 亿美元）来支持研究和教育以及改进英国的数据基础设施。	**2017 年 7 月**：《新一代人工智能发展规划》提出了直到 2030 年的人工智能发展战略。中国计划让自己的 AI 技术在 2020 年比肩世界最先进水平。到 2025 年，中国将在 AI 理论方面取得重大突破，AI 将会变成产业升级和经济结构调整的推动力。到 2030 年，中国将成为世界主要的 AI 创新中心。

达到人类水平的里程碑

人工智能达到人类水平的里程碑

在 2017 年的起始人工智能指数报告中，我们列出了 AI 达到或超越人类水平的事件时间线。这份列表列出了游戏博弈方面的成果、准确的医疗诊断和 AI 能以人类水平或超人类水平执行的常见却不复杂的其它人类任务。今年，我们为该列表增添了 4 项新成就。需要重点指出，我们不应该过度解读这些结果。下面的任务都是高度特定的，这些成就虽然很亮眼，但并不代表这些系统能将这样的能力泛化到其它任务上。

1980

奥赛罗棋（Othello）

在 1980 年代，李开复和 Sanjoy Mahajan 设计了一款基于贝叶斯学习的系统 BILL，它可以玩一个叫奥塞罗的棋牌游戏。1989 年，该系统在所有计算机玩家中获得了美国全国锦标赛冠军，并以 56-8 击败了当时位列得分榜首的美国玩家 Brain Rose。1997 年，一个名为 Logistello 的程序在与奥赛罗世界冠军保持者的对战中以 6 场全胜的战绩完胜对手。

1995

西洋跳棋

1952 年，Arthur Samuels 设计了一系列玩跳棋游戏的程序，并通过自我博弈来改进程序本身。但直到 1995 年，一个名为 Chinook 的跳棋程序才击败了当时的世界冠军。

1997

国际象棋

一些计算机科学家在 1950 年代预测计算机在 1967 年之前便可以击败人类国际象棋冠军，但直到 1997 年，IBM 公司的深蓝系统才击败了世界冠军卡斯帕罗夫。现在，即使是在手机上运行的国际象棋程序都可以达到大师水平。

2011

Jeopardy!

2011 年，IBM 公司研发的 Watson 计算机系统在著名电视智力游戏 Jeopardy! 上击败了卫冕冠军 Brad Rutter 和 Ken Jenning 并获得了一百万美元的冠军奖金。

2015

Atari 游戏

2015 年，谷歌 DeepMind 的一个团队利用强化学习系统来学习玩 49 款 Atari 游戏。系统在大部分 游戏中达到了人类水平（比如“打砖块”），然而在如“蒙特祖玛的复仇”游戏中，系统的水平还远不及人类。

人工智能达到人类水平的里程碑

2016

ImageNet 目标检测

从 2010 年到 2016 年，人工智能系统对 ImageNet 数据集图像的自动标注错误率从 28% 降到了 3%。相比之下，人类的错误率为 5%。

2016

围棋

2016 年 3 月，由谷歌 DeepMind 团队研发的 AlphaGo 系统 4-1 击败了世界围棋冠军、韩国棋手李世石。随后在 2017 年 3 月，DeepMind 发布了 AlphaGo Master，它战胜了排名第一的棋手柯洁。 2017 年 10 月，《自然》杂志的一篇论文详细描述了迄今为止的另一款新一代围棋程序 AlphaGo Zero，它以 100-0 完胜先前的 AlphaGo 系统。

2017

皮肤癌分类

Esteva 等人在《自然》杂志 2017 年发表的一篇论文中描述了一款人工智能系统，研究者训练该系统学习识别一个包含 2032 种不同疾病的 129 450 张临床图像的数据集，并比较该系统和 21 位执业皮肤科医生的诊断能力。结果显示，人工智能系统在皮肤癌诊断上的能力与专业医生不相上下。

2017

Switchboard 语音识别

2017 年，微软和 IBM 两家公司均在语音识别开发上取得进展，其语音识别系统性能非常接近在有限 Switchboard 上的人类语音识别水平。

2017

扑克

2017 年 1 月，来自卡内基·梅隆大学的程序 Libratus 在 12 万手单挑无限注德州扑克比赛中击败了 4 位顶尖的人类选手，最终赢得冠军。2017 年 2 月，来自加拿大阿尔伯塔大学的程序 DeepStack 分别与 11 位专业玩家进行了超过 3000 局的比赛。结果证明，算法根据统计学作出的判断超越了专业选手。

2017

吃豆人小姐

微软收购的深度学习团队 Maluuba 创造的 AI 系统学会了在 Atari 2600 上获得《吃豆人小姐》的最高分数值 999 900。

人工智能达到人类水平的里程碑

2018

汉语-英语翻译

微软的一个机器翻译系统在将汉语新闻翻译成英语的任务上达到了人类水平的质量和准确度。该测试是在 newstest2017 上执行的，这是一个常用于机器翻译竞赛的数据集。

2018

夺旗

DeepMind 的智能体在修改版的《雷神之锤III竞技场：夺旗》（一款流行的 3D 多玩家第一人称视频游戏）中达到了人类水平。这些智能体表现出了类人的行为，比如导航、跟随和防守。不管是作为队友还是敌人，训练后的智能体的胜率都超过较强的人类玩家，并超过了多个已有的之前最佳系统。

2018

Dota 2

OpenAI Five（OpenAI 用五个神经网络构成的战队）在有限制的 Dota 2 比赛中击败了业余人类玩家战队。OpenAI Five 通过自我对抗训练了相当于 180 年的游戏时长。*（OpenAI Five 尚未超越人类，因为它被职业人类战队击败了。）*

2018

前列腺癌分级

谷歌开发的一款深度学习系统在基于前列腺切除标本的前列腺癌分级上达到了 70% 的总体准确率。美国执业的普通病理医生的平均准确率为 61%。此外，在验证集上分级每个样本表现最好的 10 位普通病理医生中，该深度学习系统比其中 8 位更准确。

有何缺失？

有何缺失？

人工智能指数仍在发展进步，在它能真正体现 AI 的发展现状之前，还应当加入很多元素。这需要 AI 社区持续不断地输入和合作才能实现。人工智能指数应当如何延展？下面给出了三个来自 AI 专家的建议。

FRANCESCA ROSSI
IBM，帕多瓦大学（*休假中*）

常识推理和自然语言理解
AI 驱动的问答系统非常有用，但它们缺乏维持对话的能力。深度的自然语言理解仍然是一大难题，常识推理也是如此。能够衡量这些能力的指标将能帮助推动基于对话的 AI 系统的发展进步，同时也能助益 AI 的很多其它方面。人工智能指数可以添加与这些能力相关的测试和挑战进展，比如 Winograd 模式挑战赛、AI2 的 Aristo 项目和相关的科学考试 Kaggle 竞赛、IBM Debater 等基于对话的挑战赛

与人类合作
如果我们想要 AI 增强人类智能，而不是取代人类智能，那么我们不仅要关注人与机器的对比，关注“人+机器”环境也很重要。展示这方面的进展的指标可能包括人类是最终决策者但仍需机器支持的非自动系统。

AI 推理+学习
机器学习能够发现巨量数据之中的隐含关系，因此在很多应用中都非常成功。从而能够实现非常准确的预测。但是，在很多领域，我们不仅需要关系，还需要推导因果信息，以便从预测走向干预，进而塑造未来。一个可能的相关指标是 AI 会议（不只是 ML 会议，而且还有 AAAI 或 IJCAI 等 AI 会议）上有关因果推理的论文的数量。

为此，挖掘论文的标题就足够了。将 ML 与符号 AI 相结合的研究也值得关注。“ML+符号 AI”组合的论文的数量是一个可能的指标。

有何缺失？（续）

RODNEY BROOKS
麻省理工学院

带有 AI 组件的机器人
我想要谈谈机器人出货量指标。机器人出货量指标的相关性取决于来源。很多来源都包含只有很少甚至没有 AI 的工业机器人的出货量，这是一个衡量 AI 进展的糟糕指标。Rethink Robotics是我所知的一个例外。

人们还应该看看有 AI 组件的机器人，比如无人机（使用了 SLAM 和其它 AI 算法）。Roomba 等家用机器人也有 AI 组件。Roomba 应该值得关注，因为它是出货数值上最多的机器人。尽管社交型家用机器人最近出现了一些失败，但这一领域的任何成功都将极大依赖于人工智能。

TOBY WALSH
新南威尔士大学和柏林工业大学

政府开支和军事应用
过去两年最重要的 AI 进展之一可能是政府和产业界都宣布将加大对 AI 的投资力度。比如，我们已经看到，英国政府承诺投入 10 亿英镑，法国承诺投入 15 亿欧元，德国是 30 亿欧元。同时，我们也看到阿里巴巴公司宣布了 150 亿美元的投资计划，软银的 1000 亿美元愿景基金（Vision Fund）中很大一部分也会投入 AI。

人工智能指数中衡量投资增长的量化指标是有益的。这通常被描述为是一场“人工智能军备竞赛”，比如中国等国家已宣布要通过使用 AI 来实现经济和军事优势。为了衡量 AI 研究和发展重心的变迁，这样的指标可能对分析国家投资情况有用。另一个日益让人担忧的领域是军方对 AI 的应用。在某些领域，应用 AI 是受人欢迎的，比如清理地雷。但是，还存在全自动武器开发等其它一些领域，AI 界的很多人已经对此表达了严重担忧。如果能够衡量这些发展情况以及世界各地对 AI 技术的军事应用，对人工智能指数会很有助益。

致谢

致谢

人工智能指数研究是“人工智能百年研究（AI100/One Hundred Year Study on AI）”孕育出的项目。人工智能指数研究是 AI100 下的一个独立项目，现由斯坦福大学的以人为本人工智能研究院（HAI）管理。我们期待与 AI100 和 HAI 继续保持蓬勃发展的关系。

感谢谷歌、微软、字节跳动（今日头条）和 AI Journal 提供的额外的启动资金。

感谢“有何缺失？”的建议者，他们的评价提供了人工智能指数未来内容增添的信息：

Francesca Rossi, Rodney Brooks, Toby Walsh

感谢以下人士为人工智能指数2018年度报告提供了特定指标的评论：

Paul Ginsparg, WiML 理事会, Michael Chui, Barry Haddow, Sam Bowman, AAAI, James A. Landay

特别感谢 Elsevier 为我们提供了几个发人深省的数据集和可视化结果。感谢 Elsevier 的以下人士一路上提供的支持：

Ann Gabriel, Clive Bastin, Sarah Huggett, Mark Siebert, Jorg Hellwig

感谢 McKinsey & Company 和麦肯锡全球研究院，他们为我们提供了对 AI 采用情况的调查结果，并帮助分析了政府的 AI/ML 体积情况。麦肯锡的以下人士为我们提供宝贵的支持：

Heather Hanselman, Daniella Seiler, Jake Silberg

感谢以下人士不仅为我们提供数据和/或解释指标，还提供了更多帮助：

Susan Woodward, Tully Foote, Sam Davis, Lars Kotthoff, Konstantin Savenkov, Anita Huang, Calvin LeGassick

感谢 Ian Danforth （技术主管）和 Michael Chang（平面设计和封面艺术），他们是人工智能指数2018年度报告日常撰写不可或缺的一部分。

致谢

我们还要感谢下列组织、大学和会议为人工智能指数2018年度报告提供了至关重要的数据：

公司/组织

arXiv，ROS.org，Sand Hill Econometrics，Monster.com，amplified，McKinsey & Company，麦肯锡全球研究院，Prattle，IFR.org，TrendKite，艾伦人工智能研究所，欧盟委员会联合研究中心

大学

加州大学伯克利分校，斯坦福大学，伊利诺伊大学厄巴纳-香槟分校，华盛顿大学西雅图分校，卡内基·梅隆大学

清华大学，墨西哥国立天体物理学、光学与电子学院，英属哥伦比亚大学，多伦多大学，爱丁堡大学，中国科学技术大学，上海交通大学，智利天主教大学，维也纳技术大学，希伯来大学，洛桑联邦理工学院，蒙特利尔算法学习研究所

会议和倡议

AAAI, AAMAS, AI4ALL, ACL, CVPR, ICAPS, ICLR, ICML, ICRA, IJCAI, KR, NeurIPS, UAI WiML

附录

附录 1
Scopus 上的论文

返回按主题统计的 Scopus 论文
查看基础数据

数据来源
Elsevier 的 Scopus 学术出版物数据库，其中索引了近 7000 万份文档。
查看有关 Scopus 的更多信息。

方法
需要指出这种方法和前一页描述的方法的不同。这里我们使用了一种自上而下的方法——使用 Scopus 的查询工具来获取标注有关键词“Artificial Intelligence”和“Computer science”的论文的数量。这项工作由人工智能指数团队完成。

Scopus 会用关键词、发布日期、所属国家和一些其它指标来标注论文。对于按年份和地区来统计的 AI 论文，我们使用了“opensearch:totalResults”工具（会返回论文数量）来查询每个年份/国家组合的情况，如下所示：

按国家查询 AI
title-abs-key(artificial intelligence) AND PUBYEAR = {} AND AFFILCOUNTRY({})

对于每年的按主题统计的论文情况，使用了以下查询：

AI 查询
title-abs-key(artificial intelligence) AND PUBYEAR AFT {} AND PUBYEAR BEF {}
CS 查询
SUBJAREA(COMP) AND PUBYEAR AFT {} AND PUBYEAR BEF {}
全 Scopus 查询
PUBYEAR AFT {} AND PUBYEAR BEF {}

我们对 1996-2016 年期间每一年的情况都进行了上述所有查询。

细节说明
- Scopus 系统会追溯式地更新。因此，一个给定查询的论文数量可能会随时间增大。
- Elsevier 团队的成员评论说 1995 年之后的论文发表数据是最可靠的（因为 1996 年是 Scopus 数据库的起始年）。
- Scopus 论文数量包含各种质量以及与 AI 有各种相关度的论文。

对按地区统计的 AI 论文的细节说明
- 标注有多个地区的论文会重复统计。因此给定年份每个地区论文的数量之和可能不等于这一年的实际论文总数。
- “其它”包含 Scopus 上发表过 AI 论文的所有其它国家。

对按主题统计的 AI 论文的细节说明
- 2017 年的人工智能指数仅展示了 CS 类别的 AI 论文。今年，所有标注有 AI 的论文都包含了进来，不管它们是否属于更大范围的 CS 类别。
- 因为 Scopus 上的 AI 论文通常都属于 CS 类别，因此 CS 论文的增长很大程度上包含了 AI 论文的增长。
- “所有论文”类别包含所有 AI 和所有 CS 论文。

附录 2

处理 Elsevier 的 Scopus 论文数据的方法

返回 Elsevier 论文部分
查看基础数据

数据来源

Elsevier 的 Scopus 学术刊物数据库索引了接近 7000 万份文档。该数据由 Elsevier 编制。访问对应链接了解有关 Scopus 和 Elsevier 的更多信息。

关于论文索引、所属关系、地理位置和标题的深度方法可访问 Scopus 内容覆盖指南。

方法

需要重点指出这种方法和前一页描述的方法存在差异。这里我们会概述 Elsevier 统计 AI 论文数量的方法，这是一种自下而上的方法，用到了大约 800 个关键词。更多细节请参阅 Elsevier 近期的 AI 报告。

定义 AI

Elsevier 确立了一组大约 800 个与 AI 相关的关键词。然后使用这些关键词识别出了 Scopus 的 600 000 篇 AI 论文，这些论文构成了 Elsevier 的 AI 报告的基础。这些关键词是按如下方式确定的：

首先，Elsevier 使用其 Fingerprint Engine 分析了 AI 书籍、课程大纲、专利和新闻文章。Fingerprint Engine 处理了这些文档，输出了一组最相关的加权的关键词。这得到了 20 000 个相关术语。在人工审阅了这些术语后，他们选出了 797 个不同的关键词。预计 Elsevier 将在未来几个月公布这些关键词。这里有一个样本列表。

尽管如此，包含这 797 个 AI 关键词中一个或多个词的部分论文实际上并不是 AI 论文。这会产生误报——比如术语“伦理价值”具有很宽泛的范围，可能会被包含在多种非 AI 论文中。类似地，术语“神经网络”也可能属于非 AI 领域，比如生物学。

为了解决误报问题，Elsevier 引入了专家引导，通过人工选择 1500 篇关于核心 AI 和应用 AI 的 AI 文献构建了一个训练集。他们使用这个训练集训练了一个消除误报的监督式分类器，以便能处理 Scopus 的超过 600 000 篇文献的大规模数据集。

整体而言，Elsevier 的方法是为广泛了解 AI 领域情况而设计的，具有多个视角和关键词，同时还结合了监督式分类器的精度。

所属国家

作者的所属国家根据其所在的主要组织确定，这由论文的作者提供。全球性组织将默认使用其总部所在的国家，除非作者在其组织描述中指明了自己的国家。比如，认为以“Google”作为自己的组织的作者属于美国，认为以“Google Zurich”作为自己的组织的作者属于欧洲。

如果一篇论文的作者来自多个不同地区，则该论文会被重复统计。比如，如果一篇论文有来自哈佛大学和牛津大学的作者，则会被算入美国一次，再算入欧洲一次。

欧洲的定义范围是 EU44。

这些规则也适用于行业所属情况。

附录 2
处理 Elsevier 的 Scopus 论文数据的方法（续）

AI 子类别（见图）
你能在下面找到 Elsevier 确定 Scopus 上 AI 子类别的方法的总结。有关 Elsevier 方法的更多信息，参见 Elsevier 近期的 AI 报告。

为了确定广泛的子类别，Elsevier 使用了一种名为 Louvain 方法的无监督聚类技术。该方法可基于文档中的共现情况将关键词映射成聚类并展示出它们之间的联系。共现是指这些聚类并不单独出现，但彼此之间有很强的联系，比如计算机视觉文档中的神经网络。他们发现 AI 领域似乎聚类在以下区域周围：

概率推理、神经网络、计算机视觉、搜索和优化、自然语言处理和知识表征、模糊逻辑、规划和决策

Elsevier 人工智能资源中心已经发布了这些聚类的交互式图表，用户可以按地区和时间浏览各个连接和聚类。

相对活动重心（见图）
论文由专业索引器分配了主题领域，其根据论文的范围和内容分类了每篇论文的主题领域。主题领域之间并不互斥。

细节说明

- 整个数据收集过程由 Elsevier 内部完成——人工智能指数团队没有参与关键词选择或统计相关论文数量的过程。
- 这些数据接近人工智能指数2017年度报告（使用了 Scopus 但方法不同）中对应数据的两倍，而大约只有《中国人工智能发展报告》对应数据的一半，该报告基于 Web of Science的数据。
- AI 领域的范围很难界定，部分原因是很多相关领域都在快速发展，比如语音识别、计算机视觉、机器人、网络安全、生物信息学和医疗保健。限制条件也很难定义，因为 AI 的方法与很多领域有关联，比如逻辑学、概率和统计、优化、摄影测量学、神经科学和博弈论。鉴于 AI 社区对 AI 文献计量学的兴趣，我们相信如果得到研究结果的团队能够提升他们的方法的透明性，支持结果的再现性，将会非常有价值，尤其是使用了不同的文献数据库时。

附录 3

arXiv 上的论文

返回 arXiv 论文部分
查看基础数据

数据来源

arXiv.org 是一个科研文献的在线档案库，涉及物理学、数学、计算机科学、量化生物学、量化金融、统计学、电气工程和系统科学、经济学等领域。arXiv 由康奈尔大学拥有和运营。查看有关 arXiv.org 的更多信息。

方法

我们分析所使用的原始数据由 arXiv.org 的代表提供。我们选择的关键词和对应的类别如下：

人工智能/Artificial intelligence (cs.AI)
计算和语言/Computation and language (cs.CL)
计算机视觉和模式识别/Computer vision and pattern recognition (cs.CV)
机器学习/Machine learning (cs.LG)
神经和进化计算/Neural and evolutionary computing (cs.NE)
机器人学/Robotics (cs.RO)
统计学领域的机器学习/Machine learning in stats (stats.ML)

对于大多数类别，arXiv 提供了 1999-2017 年的数据。对于我们的分析，我们决定从 2010 年开始，以便将之前没有的统计学领域的机器学习包含在内。

要查看 arXiv 上其它类别的提交率，参见 arXiv.org 的提交统计数据。

细节说明

- 类别是由作者自行认定的——选择展示出来的是“主”类别。因此，需要指出并不存在简化的分类过程。此外，“人工智能”和“机器学习”类别也可通过其它子类别/关键词来分类。
- arXiv 团队成员分享说，arXiv 上的参与可能会催生更多参与——也就是说 arXiv 上一个子类别的增长可能导致特定社区看到过度索引的参与情况。
- arXiv 上的论文增长并不能反映该主题的实际论文增长情况。某些增长的原因可能是 arXiv.org 提升论文数量的工作努力，也可能是 AI 社区中研究传播的重要性越来越大。

附录 4

按国家统计的 AAAI 论文

返回按国家统计的 AAAI 论文
查看基础数据

数据来源

人工智能促进协会（AAAI）每年都会举办一些会议，其中包括年度的 AAAI 会议。2018 年按国家统计的 AAAI 论文提交/接收情况的原始数据由 AAAI 代表提供。了解更多有关 AAAI 会议的信息。

方法

我们从 AAAI 团队收集了按国家统计的 AAAI 论文提交/接收数据。AAAI 是唯一一个我们能做到这种详细程度的会议。人工智能指数报告希望在未来的报告中囊括其它会议的同等数据。

细节说明

- 本分析中包含的国家都至少向 AAAI 会议提交了 10 篇论文。
- 这些数据来自 2018 年的会议，该会议于 2018 年 2 月在路易斯安那州新奥尔良举办。其它年份的提交/接收论文图景可能看起来不一样。
- 由于空间限制，接收率很大程度上是有限的，下面引述了 AAAI 对 AAAI 2019 会议的空间限制的论述（1/27/19-2/1/19，夏威夷州火奴鲁鲁）：

  > “今年（2019 年）我们收到的提交数量是创纪录的，超过 7700 篇。其中 7095 篇得到了评议，而由于空间限制，我们仅能接受 1150 篇论文，接受率为 16.2%。”——AAAI

附录 5

美国的 AI 和 ML 课程注册情况

返回美国课程注册情况
查看基础数据

数据来源

课程注册数据是直接从每所大学收集的。学生总数收集自学校档案（通常来自注册办公室网站）。我们的分析包含了以下大学：

加州大学伯克利分校，斯坦福大学，伊利诺伊大学厄巴纳-香槟分校，华盛顿大学

方法

我们联系了计算机科学领先的大学以了解入门级 AI 课程和入门级 ML 课程的学生注册人数随时间的变化情况。几所学校参与了这项调查。如果一所学校没有足够的历史数据或数据过小，那么我们的分析中就不会包含其注册数量。

有些学校的注册人数是按学期统计的，有的则是按年提供。为了比较这些学校，我们将按学期统计的注册数据进行了聚合，以展示整个学年的情况。此外，某些学校有多个课程都可被视为“入门级”，而其它一些学校仅有一个课程。在合适且相关的情况下，我们将多个课程组合到了一起，以呈现“入门级 AI”的趋势线。

对于将注册情况以占本科生群体的百分比进行呈现，每年的 AI/ML 注册人数除以当年的本科生总数。这样计算是为了更为公平地比较各校的注册趋势。

细节说明

- 几乎每所学校都强调，注册人数（尤其是近些年的注册人数）与课程供应相关，而不是学生的需求。我们的数据展示了成功注册课程的人数，但没有考虑等待名单或其它需求指标。
- 这些课程通常只向本科生开放，而且通常专业学生和非专业学生都可选修。
- 某些课程随时间而修改了名称，下面给出了 2017 年的课程名称。我们还列出了学校主管部门给出的有关注册数据的细节说明。学校主管部门没有提到的任何细节也没有在下面给出。

AI 课程：
Berkeley CS 188
Stanford CS 221
UIUC CS 440
- UIUC 的代表将增长归因于更大的教室/其它因素，从而能满足学生数量增长的需求。

UW CSE 415 / 416 (非专业) & CSE 473 (专业)
- CSE 416 是 2017 学年新开的课程，为 2017 年的 AI 课程注册人数贡献了一些增长。

ML 课程：
Berkeley CS 189
Stanford CS 229
UIUC CS 446
UW CSE 446

附录 6

国际的课程注册情况

返回国际的课程注册情况
查看基础数据

数据来源

课程注册数据是直接从每所大学收集的。我们的分析中包含以下大学：

清华大学（中国），国家天体物理学、光学与电子研究院（墨西哥），不列颠哥伦比亚大学（加拿大），多伦多大学（加拿大），爱丁堡大学（苏格兰），中国科学技术大学（中国），上海交通大学（中国），智利天主教大学（智利），维也纳技术大学（奥地利），希伯来大学（以色列），洛桑联邦理工学院（瑞士），蒙特利尔算法学习研究所（加拿大魁北克）

方法

参见 附录 5 的方法。

细节说明

- 几乎每所学校都强调，注册人数（尤其是近些年的注册人数）与课程供应相关，而不是学生的需求。我们的数据展示了成功注册课程的人数，但没有考虑等待名单或其它需求指标。
- 不同于研究中的美国学校，其它国家的学校在课程是否仅对本科生开放方面差异很大。
- 图 1 给出了 AI 和 ML 课程总共的增长。图 2 仅展示了 AI 课程的注册情况。我们这样做是为了在每张图中都展示相似的数据。在某些案例中，我们得到了一些学校的其它数据，但没有展示出来，因为我们希望各所学校都有同类信息进行比较。其它数据可访问右上角的基础数据链接。
- 某些课程随时间而修改了名称，下面给出了 2017 年的课程名称。我们还列出了学校主管部门给出的有关注册数据的细节说明。学校主管部门没有提到的任何细节也没有在下面给出。

国家天体物理学、光学与电子研究院（INAOE，墨西哥）- 课程： C141 (AI) 和 C142 (计算学习)
注：INAOE AI/ML课程注册人数受到了 INAOE 研究生项目整体录取学生人数的很大影响。INAOE 的代表说 INAOE 学生总数也有增长，因此也会影响 AI/ML 课程的注册人数。
中国科学技术大学（中国） - 课程：中国科学技术大学有多个系都有入门级 AI/ML 课程，包括计算机科学与技术系、自动化系、信息科学与技术系和数据科学系。
爱丁堡大学（苏格兰） - 课程：应用机器学习入门（本科生和研究生）和信息学 2D——推理和智能体（仅本科生）
上海交通大学（中国） - 课程：CS410（人工智能本科入门）
智利天主教大学（智利） - 课程：人工智能入门
在 2017 年之前，该课程每学期一次。相对于 2018 年，2017 年的巨大需求是因为从一门课程变成了两门。
清华大学（中国） - 课程： AI (60240052 & 40250182) 和 ML (00240332 & 70240403 & 80245013 & 80250943)
对本科生和研究生开放。
多伦多大学（加拿大） - 课程： AI (CSC384) 和 ML:(CSC411)
2016 年是夏季 AI 课程开放的第一年。由于需求增长，他们在 2015 年决定开放两学期的 ML 课程。
不列颠哥伦比亚大学（加拿大） - 课程： AI 和 ML (CPSC 322, CPSC 340, CPSC 422)
维也纳技术大学（奥地利） - 课程： AI 入门（本科生）和 ML (研究生)
对 CS 学生而言，这两门课程都不是必修课。
希伯来大学（以色列） - 课程： AI 入门 (67842) 和 ML 入门 (67577)
洛桑联邦理工学院（EPFL，瑞士） - 课程： 人工智能（本科生）和智能体（研究生）
EPFL 的代表告诉我们由于时间安排和资源，每年都有很大变化。
蒙特利尔算法学习研究所（MILA，加拿大） - 课程：机器学习（仅限研究生，非高级）
MILA 是只针对研究生的项目。MILA 的代表指出该 ML 课程不是为核心 ML 研究生设立的，而是为其它专业的想学习 ML 的学生提供的。核心 ML 的学生有更高级的课程。2014 年没有课程。

附录 7

师资力量多样化

返回师资力量多样化
查看基础数据

数据来源

师资力量多样性是以人工方式在 2018 年 9 月 21 日各学校的 AI 系网站上收集的。展示的学校都是能够访问 AI 师资力量名册的计算机科学领先的大学。

方法

为了得到 AI 师资力量性别多样性的细节情况，教授姓名是在学校网站上收集的，然后使用了姓名和照片来识别性别。下面可访问每所学校的更多具体细节：

加州大学伯克利分校 – 参阅教职人员链接
包括助理教授、主要和次要的教职人员
斯坦福大学 - 参阅教职人员链接
包括教职人员和研究科学家以及附属教职人员
伊利诺伊大学厄本那-香槟分校 - 参阅教职人员链接
包括 CS 教职人员和附属教职人员
卡内基·梅隆大学 - 参阅教职人员链接
包含所有列出的教职人员
伦敦大学学院- 参阅教职人员链接
包含 People 链接下的所有教职人员
牛津大学 - 参阅教职人员链接
仅包含 *Faculty* 部分的教职人员
苏黎世联邦理工学院 - 参阅教职人员链接
仅包含有“博士”头衔的教职人员

细节说明

- 学校数据是在 2018 年 9 月 21 日取出的。在本年度的人工智能指数报告发布时，学校教职人员可能已发生变化。
- 数据取自学校的 AI 教职人员名册，并没有算入访问教授或注册在其它科系的教授。类似地，如果一位教授是活跃的 AI 教职人员，就算该教授属于另一个系，也会算入这位教授。
- 我们使用了教授的姓名和照片来分配性别——这样的操作可能会让一些教授的性别识别出错。我们的意图完全不是对性别做出错误的预设，而是展示当前 AI 学术界女性代表性的现状。
- 学术界的女性代表性并不等同于产业界中的女性代表性（事实上，产业界中的女性比例通常显著更低）。
- 这些图片展示了当前的代表性，并没有考虑随时间的改善情况，参阅斯坦福大学教职人员 James A. Landay 博士的以下声明：

> “我们非常重视聘请更多样性的教职人员。这份名单上大部分女性都是在过去 2-3 年里聘请的，所以我们一直都在进步。”——Landay 博士，斯坦福大学

附录 8
参会情况

返回参会情况
查看基础数据

数据来源

参会数据是从会议/组织代表那里直接收集的。我们收集了以下会议的数据：
AAAI -人工智能促进协会
AAMAS – 国际自动化智能体和多智能体系统会议
AI4ALL
ACL - 计算语言学协会
CVPR – 计算机视觉和模式识别会议
ICAPS - 国际自动规划和调度会议
ICLR - 国际学习表征会议
ICML – 国际机器学习会议
ICRA – 国际机器人与自动化会议
IJCAI – 国际人工智能联合会议
KR – 国际知识表征和推理原理会议
NeurIPS - 神经信息处理系统会议
UAI – 人工智能不确定性会议
WiML – 机器学习中的女性

方法

我们将 2018 年参会人数至少 2000 的会议定义为“大型会议”，将 2018 年参会人数少于 2000 的会议定义为“小型会议”。所选择的会议都是引领 AI 研究并能提供年度参会数据的会议。

选择 AI4ALL 和 WiML 的原因是它们在 AI 包容性和数据可用性方面的进展。我们希望能在未来的报告中添加更多面向代表性不足的群体的组织/会议。

细节说明

有关会议的细节说明

- 某些会议组织者仅能提供估计的参会人数——我们将其估计数量视为准确值。
- 某些会议并不是每年都举办，有的跳过了一些年份。
- 有几位会议组织者告诉我们，因为会议场地要提前一年确定，所以场地供应往往有限。因此，参数人数并不一定能反映需求。

有关 AI4ALL/ WiML 的细节说明

- 需要重点指出，也存在一些其它的支持 AI 领域包容性的正式和非正式项目。
- 参与情况并不代表该领域在纳入女性和代表性不足的群体方面所取得的进展。

附录 9

机器人软件下载情况

返回机器人软件下载情况
查看基础数据

数据来源

ROS.org 网页浏览量和机器人软件下载数据来自 ROS.org 社区指标报告。参阅每年数据的 PDF 文档。了解更多有关 ROS.org 的信息。

细节说明

- ROS.org 页面浏览量并不代表总页面浏览情况，仅能代表顶级国家的页面浏览情况。90 多个更低体量的国家不包含在内。
- ROS.org 上提供了 2018 年的指标，但未被人工智能指数纳入，因为他们使用了稍有不同的方法。在这里了解更多。
- 月平均值是使用每年的 7 月计算的（只有 2012 和 2013 年分别使用了 6 月和 8 月）。
- ROS.org 的代表指出其机器人软件在非英语国家的活动可能被低估了，因为这些指标仅考虑了英语网站。
- ROS.org 的代表推测 2015 年的下降很可能是因为采样问题，而且那个月发布的补丁也更少，这反映在了下载数据上。在 这里 查看考虑了这一点的指标。

附录 10
创业公司/投资

返回创业公司/投资

无法提供基础数据

数据来源

使用 Dow Jones VentureSource 数据库的 Sand Hill Econometrics 提供了创业公司和风险投资数据。使用了 Crunchbase API 来证实 Sand Hill 数据库中的 AI 公司名单。更多有关 Sand Hill、VentureSource 和 Crunchbase 的信息请访问对应链接。

方法

我们使用了下列关键词来获取 Crunchbase 上的 AI 相关公司名单：

类别名称	Crunchbase 类别 UUID
Artificial Intelligence（人工智能）	c4d8caf35fe7359bf9f22d708378e4ee
Machine Learning（机器学习）	5ea0cdb7c9a647fc50f8c9b0fac04863
Predictive Analytics（预测分析）	ca8390d722c65bb5f87022f52f364b1b
Intelligent Systems（智能系统）	186d333a99df4a4a6a0f69bd2c0d0bba
Natural Language Processing（自然语言处理）	789bbbefc46e1532a68df17da87090ea
Computer Vision（计算机视觉）	a69e7c2b5a12d999e85ea0da5f05b3d3
Facial Recognition（人脸识别）	0b8c790f03bcb2aba02e4be855952d6d
Image Recognition（图像识别）	af9307c9641372aeaac74391df240dd2
Semantic Search（语义搜索）	bb1777e525f3b9f2922b0d5defe0d5bb
Semantic Web（语义网）	ac34c4c6e430f66f44aeef8ca45b52bb
Speech Recognition（语音识别）	24dad27e6a49ccc1ec31854b73d60d16
Text Analytics（文本分析）	e7acbc9932c74c8809ed8fae63eb2559
Virtual Assistant（虚拟助手）	2eb0b56b6896e0f0545fc747a9ba6751
Visual Search（视觉搜索）	ad8f33a8c5d0b7d786f389aa2954c119

Crunchbase 返回了 4477 家 AI 公司。然后这份机构列表与 VentureSource 数据库中的所有获得了风险投资的公司的列表进行了交叉参照——来自Crunchbase 的 4477 家公司中有 1096 家匹配上了 VentureSource 列表中的 AI 公司。Sand Hill 对匹配后的列表进行了汇总，给出了随时间变化的公司数量和投资数额。

“所有创业公司”的数据是仅使用 VentureSource 数据库收集的，不涉及 Crunchbase API。

细节说明

- 尽管我们发现风险投资趋势也值得分享，但 VentureSource 数据是专有的，我们很遗憾不能向公众公开这些数据。
- 所有与 VentureSource 产品的互动都是由 Sand Hill Econometrics 执行的。
- 公司一旦上市、被收购或倒闭，就会被移出 VentureSource 数据库，但是该公司在被移除前的融资轮次仍会留存在该数据库中。
- Crunchbase 和 VentureSource 都难以识别开发 AI 系统的公司和仅仅部署使用 AI 系统的公司。我们内部分别称它们为 AI“制造者”和“使用者”。我们随机测试了匹配公司中的 109 个样本（占匹配结果的 10%），发现其中有 68 家公司是 AI“制造者”（62%）。其余的 38% 虽然仍被视为 AI 公司，但可能本身并不开发 AI 系统。
- 我们的 Crunchbase AI 类别列表目前仅支持机器学习技术。
- 活跃的 AI 创业公司数量仅使用每年 1 月的数据。

附录 11
工作就业

返回工作就业
查看基础数据

数据来源
有关 AI 技能工作需求和 AI 工作申请者性别多样性的图都是由求职网站 Monster.com 提供的。Monster.com 使用了 Garner TalentNeuron 来获得结果，这是一个在线人才市场情报门户网站。更多信息请访问 Garner TalentNeuron 或 Monster.com。

大型科技公司工作需求的图由 Indeed.com 直接提供。更多信息请访问 Indeed。

方法
Monster.com
Monster.com 使用了 Garner TalentNeuron 数据库来搜索以下项。如果该项列在工作岗位的所需技能列表中，则就会包含该工作。在搜索时，每一项还都会与关键词“人工智能”进行配对。比如“机器人学”只有当工作描述中包含了“人工智能”和“机器人学”两者时才会纳入考量。来自 Monster.com 的工作仅限于美国。

人工智能+
自然语言处理，机器学习技术、深度学习、计算机视觉、语音识别、机器人学

Indeed
来自 Indeed.com 的数据使用了来自亚马逊、谷歌、Facebook、IBM、微软和苹果的全球 AI 工作岗位招聘数据。如果一个工作招聘中包含关键词“机器学习”、“人工智能”、“自然语言处理”和/或“深度学习”，则就将该工作包含在内。然后每家公司 AI 相关岗位需求会被转换成占总工作岗位需求的比例。比例按月份进行了分组，给出了从 2013 年 10 月 1 日到 2018 年 10 月 1 日的每三个月的平均需求变化情况（每个月的值都是该月与之前两个月的平均）。然后该数据集根据 2013 年 10 月的数据进行了归一化。

科技公司必须结合起来才能达到一定的可信度。

细节说明
- 包含超过一个 AI 相关技能的工作会被重复统计；某些技能还意味着需要上面没有列出的某些技能。

对 AI 工作申请者的性别多样性的细节说明
- 申请并不意味着会聘用，也不能代表整个行业。
- 本报告仅考虑了“男性”和“女性”性别。

附录 12
专利

返回专利
查看基础数据

数据来源
专利数据由 Amplified 提供，其提供专利搜索服务。

方法
所有的方法和基础数据参考 Amplified 的详细说明。

细节说明

- 我们的图片没有展示 2015-2017 年的数据，因为这几年没有完整的专利数量统计。（这是因为申请到发布之间有很长的时间间隔）。

尽管存在一些跟踪 AI 专利随时间和跨地区的情况的尝试，但并不存在一种标准方法。因此，我们给出的专利数据可能不同于其它组织机构的数据。

这份 清华大学的中国人工智能发展报告 值得一提，该报告发布于 2018 年夏季，表明中国拥有最多的已公布专利，美国和日本紧随其后。尽管如此，我们并不清楚他们使用了什么专利代码和定义而得到了这样的结果。每个地区的实际专利数量也并不容易在图中看出。

由于该领域很难定义，因此人工智能指数报告在分析中完全透明地给出了专利代码和关键词。我们也鼓励读者指出我们的方法的缺陷，以便本报告最终能全面地反映这个领域。

附录 13
AI 采用情况调查

返回 AI 采用情况调查
查看基础数据

数据来源
该调查由 McKinsey & Company （麦肯锡）编写、填充和分析完成。你可以在 这里 找到来自麦肯锡 AI 采用情况调查的更多结果。

方法
麦肯锡对自己的方法给出了以下声明：
该调查是在该领域内线上进行的，调查时间为 2018 年 2 月 6 日至 2018 年 2 月 16 日。调查收到了 2135 位参与者的回应（共邀请了 20466 位参与），他们来自麦肯锡的"在线高管小组（Online Executive Panel）"，能代表各个地区、行业、公司规模、职能特性和职位。其中，1657 位参与者表示在调查提及的 9 类人工智能技术中，他们的公司有一个或多个职能或业务部门至少在试点其中一项人工智能技术。所有的调查参与者都是这个在线小组的成员，该组包含 McKinsey.com 超过 30000 位选择参与麦肯锡专有调研的注册用户，其中包含来自世界各地的高管、经理以及各职位层级的雇员。本次调查的样本覆盖 108 个国家；为了调整回应率的差异，受访数据根据每位回应者所在国家对全球的 GDP 的贡献进行了加权。

麦肯锡为人工智能指数报告提供了选择以下问题的每个答案选项的回应者的百分比：

问题 1：你如何描述你的组织当前对以下 AI 技术的使用情况？（机器人过程自动化、机器学习、虚拟代理或会话接口、计算机视觉、自然语言文本理解、自然语言语音理解、自然语言生成、实体机器人、自主车辆）。请选择一项：
1. 完全没使用过
2. 至少一个职能或业务部门在试点
3. 至少一个职能或业务部门将其纳入了标准业务流程中
4. 跨多个职能或业务部门将其纳入了标准业务流程中
5. 不知道

问题 2：你的组织有哪些业务职能中部署了 AI？（经过随机排列，选择所有符合条件的选项）
1. 市场营销与销售
2. 战略和公司财务（例如，资本分配决策）
3. 风险（即风险管理、欺诈和债务）
4. 人力资源
5. 产品和/或服务开发
6. 供应链管理
7. 生产制造
8. 服务运营（例如，现场服务、客户服务、后台）
9. 以上都不是
10. 不知道

对于问题 1，我们要求麦肯锡根据地理位置来划分回应情况。然后人工智能指数将答案 3 和 4 进行了求和，得到了至少有一个职能或业务部门纳入了 AI 能力的结果。对于问题 2，麦肯锡提供了按地理位置和按行业的划分结果。不同于 Q1，Q2 仅限于公司已经纳入或正在试点 AI 技术的回应者。每张图片下的注释说明了相关响应者的数量。

细节说明
- 调查回应者的答案受限于他们对自己公司的 AI 采用情况的认知。

附录 14

企业主动性：公司营收电话会议调查

返回营收电话会议调查
查看基础数据

数据来源
所有的营收电话会议调查提及数据都由 Prattle 提供，这是一家专注于文本分析和情感分析的投资研究公司，Prattle从 FactSet 获取了营收电话会议调查的转录结果。

方法
我们要求 Prattle 提供电话调查中提到了“人工智能”和“机器学习”的调查随时间变化的统计次数。我们也要求提供提到了“云”和“大数据”的统计数据，以提供更多背景和基准。结果由公司提供，包括每家公司的行业和规模。

为了从电话转录文本中取出语言，Prattle 使用了正则表达式（regrex）的变体来搜索在 AI 讨论中使用的每个字符串文本和语言模式。“人工智能”和“机器学习”的提及情况也包括它们的缩写形式“AI”和“ML”。然后 Prattle 移除了误报案例（比如“ML”也是“毫升”的缩写，因此 Prattle 移除了“ML”之前有数值的情况）。

行业分类使用了标准普尔的标准分类方式 GICS 。规模分类（本报告中没有使用，但可从基础数据中获取）基于市值分类，这是 Prattle 使用已发行股票和销售价格的数据算出的。Prattle 提供的范围如下：

```
mkt_cap <= '500000000'  THEN 'micro'::text
mkt_cap <= '4000000000' THEn  'small'::text
mkt_cap <= '20000000000' THEN 'mid'::text
mkt_cap IS NOT NULL THEN 'large'::text
else Null
```

Prattle 为人工智能指数报告提供了包含日期、公司名称、行业、公司规模和提及了“人工智能”、“机器学习”、“云”和“大数据”术语的次数的电子表格。对这些数据的所有分析由人工智能指数团队完成。

细节说明
- 提及并不一定代表该公司投资或采用了 AI。
- 本分析仅包含上市公司。

附录 15
机器人进口情况

返回机器人进口情况
查看基础数据

数据来源
数据直接取自国际机器人联合会（IFR）2014 年、2015 年和 2017 年的《世界机器人报告》。访问以下链接了解更多。更多有关 IFR 的信息。

2012 年机器人进口数据
来源：2014 年世界机器人报告-执行摘要（表 1）
2013-14 年机器人进口数据
来源：2015 年世界机器人报告-执行摘要（表 2）
2015-16 年机器人进口数据
来源：2017 年世界机器人报告-执行摘要（表 4.1）

方法
展示的数据是各个国家进口的工业机器人的数量。工业机器人的定义标准是 ISO 8373:2012 标准。

参阅有关 IFR 的方法 的更多信息。

细节说明
- 目前还不清楚如何确定运行了可被归类为“AI”的软件的机器人的数量比例，也不清楚哪种程度的 AI 发展会有助于工业机器人的使用。

附录 16
GitHub 星数

返回 GitHub
查看基础数据

数据来源
我们使用了存储在 Google BigQuery 上的 GitHub 存档。

方法
本报告中的图片展示了多个 GitHub 库随时间变化的星数。这些库包括：

apache/incubator-mxnet, BVLC/cafe, cafe2/cafe2, dmlc/mxnet, fchollet/keras, Microsoft/CNTK, pytorch/pytorch, scikit-learn/scikit-learn, tensorflow/tensorflow, Theano/Theano, Torch/Torch7

GitHub 存档数据保存在 Google BigQuery 上。我们使用了 Google BigQuery 来统计每个所涉库的 “WatchEvents”的数量。收集 2016 年阶段的数据的代码样本如下：

```
 SELECT
                project,
                YEAR(star_date) as yearly,
                MONTH(star_date) as monthly,
                SUM(daily_stars) as monthly_stars
FROM (
SELECT
                repo.name as project,
                DATE(created_at) as star_date,
                COUNT(*) as daily_stars
FROM
                TABLE_DATE_RANGE(
                                [githubarchive:day.],
                                TIMESTAMP(“20160101"),
                                TIMESTAMP(“20161231"))
WHERE
                repo.name IN (
                                “tensorflow/tensorflow",
                                 “fchollet/keras",
                                “apache/incubator-mxnet",
                                 “scikit-learn/scikit-learn",
                                "cafe2/cafe2", "pytorch/pytorch",
                                 "Microsoft/CNTK", “Theano/Theano",
                                "dmlc/mxnet", "BVLC/cafe")
                AND type = ‘WatchEvent’
GROUP BY project, star_date
)
GROUP BY project, yearly, monthly
ORDER BY project, yearly, monthly
```

细节说明见下一页

附录 16
GitHub 星数（续）

细节说明
这个 GitHub 存档目前还没提供统计用户移除某个库的星标的方式。因此，报告的数据可能略高于估计的星数。与 GitHub 上各个库的实际星数比较发现，数字相当接近而且趋势是一样的。

还存在其它检索 GitHub 星标数据的方法。我们使用了 star-history 工具来抽查我们的结果。

尽管 GitHub 项目的 Fork 情况也值得调查，但我们发现各个库的标星趋势与 Fork 趋势几乎一模一样。

附录 17
公众的兴趣

返回公众的兴趣
查看基础数据

数据来源
所有数据都是从 TrendKite 检索的。TrendKite 服务索引了一般媒体的文章并使用了一种情感分析分类器为这些文章分配了“正面”、“负面”或“中立”标签。了解 TrendKite 的更多信息。

方法
我们使用了以下查询来识别 AI 文章。我们对其进行了调整，移除了一个发布了不成比例的带有负面情绪的无关文章的来源。去年，中立文章部分是暗含的，不包含在图中。今年我们在图中包含了中立文章。

查询
”Artificial Intelligence”
AND NOT “MarketIntelligenceCenter.com’s"

NOT site_urls_ll:(
“individual.com"
OR "MarketIntelligenceCenter.com")

过滤条件
- 仅包含英语文章
- 移除新闻稿
- 移除金融新闻
- 移除讣告

附录 18
政府提及情况

返回政府提及情况
查看基础数据

不同国家的数据来源、方法和细节说明
数据的收集和分析由麦肯锡全球研究院（MGI）完成。

加拿大（下议院）：
数据是使用加拿大议会网站上的 Hansard 搜索功能收集的。MGI 搜索了术语 "Artificial Intelligence" 和 "Machine Learning" （含引号）并将结果下载为 CSV 文件。日期范围设置为“All debates”。数据截至 2018 年 11 月 20 日。网上可获取数据的最早日期是 2002 年 8 月 31 日。

每个计数都表示在下议院的议程中“人工智能”或“机器学习”在一个评论或论述中被至少提及了一次。也就是说在一个事件或对话中，如果某人在评论或论述中提到了多次“人工智能”或“机器学习”，则只计算一次。但是，如果在同一事件中发言人在不同的评论（其发言之间有其它人的发言）中都提到了“人工智能”或“机器学习”，则就会计算多次。对“人工智能”和“机器学习”的统计是分开进行的，因为执行的是不同的搜索。提到“AI”和“ML”缩写的情况不包含在内。

英国（下议院、上议院、威斯敏斯特大厅和委员会）
数据是使用英国议会的 Hansard 网站的 Find References 功能收集的。MGI 搜索了术语 "Artificial Intelligence" 和 "Machine Learning" （含引号）并对结果进行了编排。数据截至 2018 年 11 月 20 日。网上可获取的数据可追溯到 1800 年 1 月 1 日。包含根据 Open Parliament Licence v3.0 许可的议会信息。

与加拿大的数据类似，每次计数意味着“人工智能”或“机器学习”在议程的一个评论或论述中被至少提及了一次。因此，如果一位成员在其论述中多次提到“人工智能”或“机器学习”，则仅统计一次。但是，如果在同一事件中发言人在不同的评论（其发言之间有其它人的发言）中都提到了“人工智能”或“机器学习”，则就会计算多次。对“人工智能”和“机器学习”的统计是分开进行的，因为执行的是不同的搜索。提到“AI”和“ML”缩写的情况不包含在内。

美国（参议院和众议院）
数据是使用美国国会记录网站的高级搜索功能收集的。MGI 搜索了术语 "Artificial Intelligence" 和 "Machine Learning" （含引号）并将结果下载为 CSV 文件。没有选择“词变体”选项，议程包括参议院、众议院、论述扩展（Extensions of Remarks），但不包括每日摘录（Daily Digest）。数据截至 2018 年 11 月 20 日。网上可获取的数据可追溯到第 104 届国会（1995 年）。

每次计数意味着“人工智能”或“机器学习”在国会记录中的一次特定事件中被至少提及了一次，包括阅读法案。因此，如果一位发言人在其论述中多次提到“人工智能”或“机器学习”，或多位发言人在同一事件中提到了“人工智能”或“机器学习”，则仅统计一次。对“人工智能”和“机器学习”的统计是分开进行的，因为执行的是不同的搜索。提到“AI”和“ML”缩写的情况不包含在内。

附录 19

ImageNet 训练时间

返回 ImageNet
查看基础数据

数据来源

ImageNet 训练时间是通过审阅 arXiv 文献收集的，使用了以下论文：

2017 年 6 月 8 日：60 分钟
2017 年 11 月 1 日：30 分钟
2017 年 11 月 12 日：15 分钟
2018 年 7 月 30 日：6.6 分钟
2018 年 11 月 13 日：3.7 分钟

细节说明

去年我们展示了 ImageNet 大规模视觉识别挑战赛（ILSVRC）的结果，这是一个算法之间比拼大规模目标检测和图像分类的竞赛。尽管 ImageNet 竞赛已于 2017 年终止，但该数据库仍然可用。

附录 20
句法分析

返回句法分析
查看基础数据

数据来源
Penn Treebank 的“华尔街日报”部分是一个有标注的句子数据集，其中每个句子都标注了一个基于句子成分的解析树。该数据集的第 23 部分已经成为自动句法分析器研究的主要测试集。更多信息参见 Penn Treebank.

方法
自动句法分析器的评估方式是比较分析器自动生成的成分与测试集的基准句法成分。生成的成分的精度和召回率结合起来作为 F1 分数。我们报告的是句法分析器在 Penn Treebank 的 WSJ 部分的第 23 部分上的 F1 分数。我们报告了不同句子长度的分数。

我们咨询了由 nlpprogress.com 维护的英语句法成分分析排行榜。我们还进行了额外的文献审阅，以确认列表中的分数和搜索更新的最佳结果。新的数据点与人工智能指数2017年度报告的数据进行了结合。报告 F1 分数的句法分析器包含集成模型，而不只是单个模型。

细节说明
- 在自动句法分析研究早期，由于计算和方法的原因，通常是在少于 40 词和少于 100 词的句子上进行评估。我们记录了系统在少于 40 词的句子上的 F1 分数，以及在该语料库中所有可用句子上的 F1 分数。

附录 21
机器翻译（BLEU）

返回机器翻译
查看基础数据

数据来源
数据来自 EuroMatrix，提供了 Workshop on Machine Translation（WMT）竞赛的 BLEU 分数矩阵。

方法
每年 WMT 都会推出一个新闻翻译任务并会提供新的训练和测试数据集。参与团队可以提交他们构建的翻译系统来参与该翻译任务。

WMT 的主要指标的目的是评比互相竞赛的系统，而不是提供年度比较。这也非常耗费劳力。我们回到了 BLEU，这是一种粗略比较系统翻译结果和人类翻译结果的自动方法。这是一种修改版的精度，取值 0 到 1 之间，越高越好。人们也可以在翻译对语料库上计算机器翻译系统的平均 BLEU 分数。我们记录了每年提交给其新闻翻译任务的英德翻译和德英翻译系统的最高平均 BLEU 分数。

EuroMatrix 已经记录了自 2006 年以来在英德翻译和德英翻译语言对新闻翻译任务上的 BLEU 分数。我们选择了每年 BLEU 分数最高的系统的结果，具体使用了 BLEU (11b)，这定义了一个句子 token 化的协议。如果可能，我们选择在这两个语言对上 BLEU 排名都很高的系统的分数作为当年的代表。

细节说明

- BLEU 可以自动计算得到，而且已被证明与人类对翻译质量的评估有关。但是，该指标不能跨语料库使用，这会误导对不同系统的 BLEU 分数的比较。
- 尽管图上有整体向上的趋势，但可以看到 2017 年有一个缺陷——相比于 2016 年的分数有明显的下降（但是 2017 年的分数仍高于 2015 年）。2017 年的机器翻译系统不太可能比 2016 年更差，只是因为这里给出的评估方案并不完善。但是，从更大的时间跨度上看，BLEU 分数仍能代表机器翻译领域的进展。

附录 22
ARC

返回 ARC
查看基础数据

数据来源
AI2 推理挑战赛（ARC）由艾伦人工智能研究所举办。ARC 表现水平数据来自 ARC 排行榜。请在对应的链接查看简易集和挑战集的排行榜。

方法
参与者可通过艾伦人工智能研究所的网站下载 ARC 数据集和向排行榜提交结果。

来自简易语料库的问题示例：

> 哪种技术是最近期开发出来的？（A）手机（B）电视（C）冰箱（D）飞机（四年级）
>
> 一名学生假设藻类是生产者。下列哪个问题能最好地帮助该学生知道这个假设是否正确？（A）藻类是否消费其它生物？（B）哪些生物消费藻类？（C）藻类是否使用太阳光产生食物？（D）没有藻类的生态系统能否存活？（八年级）

来自挑战语料库的问题示例：

> Juan 和 LaKeisha 在斜坡上滚动几个物体。他们想知道哪个物体能滚得最远。他们应该怎么做以便能重复他们的调查？（A）将物体分组（B）改变斜坡的高度（C）选择滚动不同的物体（D）记录调查的细节（四年级）
>
> 高压系统会阻止空气上升到大气中更冷的水会冷凝的区域。如果高压系统在一个地区持续了较长时间，最可能导致下列哪种结果？（A）雾（B）降雨（C）干旱（D）龙卷风（八年级）

每个问题值 1 分。模型可以选择多个选项，在这种情况下，如果其 N 个答案中有一个是正确的，则模型得 1/N 分，否则得 0 分。整体分数是各个问题上所得分数的平均。

人工智能指数报告仅收集了超过之前提交结果的分数。如果某个提交结果低于之前的任何提交结果，就不会包含在我们的图表中。

更多有关规则和提交指南的信息参阅这里。

附录 23
GLUE

返回 GLUE
查看基础数据

数据来源
GLUE 基准数据来自 GLUE 排行榜。参见有关 GLUE 基准的更多信息。

方法
参与者可通过 GLUE 网站下载 GLUE 任务以及向排行榜提交结果。分数是基于每个任务各自的指标计算的。所有指标均会乘上 100（即以百分数的形式呈现）。然后再对这些分数求平均，得到最终分数。对于有多个指标的任务（包括 MNLI），各指标进行了平均。

排行榜上默认仅会展示分数排行靠前的提交。其它提交可在每位用户的展开视图下查看。参赛者可以私密提交，可防止结果显示出来。人工智能指数报告的图表不包含任何私密提交。为了评分，MNLI 匹配和不匹配被视为同一个任务。

人工智能指数仅收集了超过之前提交结果的分数。如果某个提交结果低于之前的任何提交结果，就不会包含在我们的图表中。

更多有关规则和提交指南的信息参阅这里。