AliResearch 阿里研究院 × 智谱·AI

2023 全球数字科技发展研究报告

PART 2：科技人才储备实力研究

数字技术前瞻

报告作者



叶静芸

jingyun.ye@aminer.cn

智谱 AI 研究咨询部科技产业研究员，生态学博士。曾经多次主持或参与“北京市十大高精尖产业研究”、“各地企业技术尽调”、“人工智能系列报告”等多个项目，积累了科技政策、人才政策、科技评估等多方面研究经验。重点关注人工智能、先进制造等领域。



邬 克

ke.wu@aminer.cn

智谱 AI 研究咨询部科技产业研究员、负责人，外交学院法学硕士。在核心期刊、知名媒体发表论文和评论性文章 20 余篇；10 余年科技产业研究咨询经验。



苏 中

suzhong.sz@alibaba-inc.com

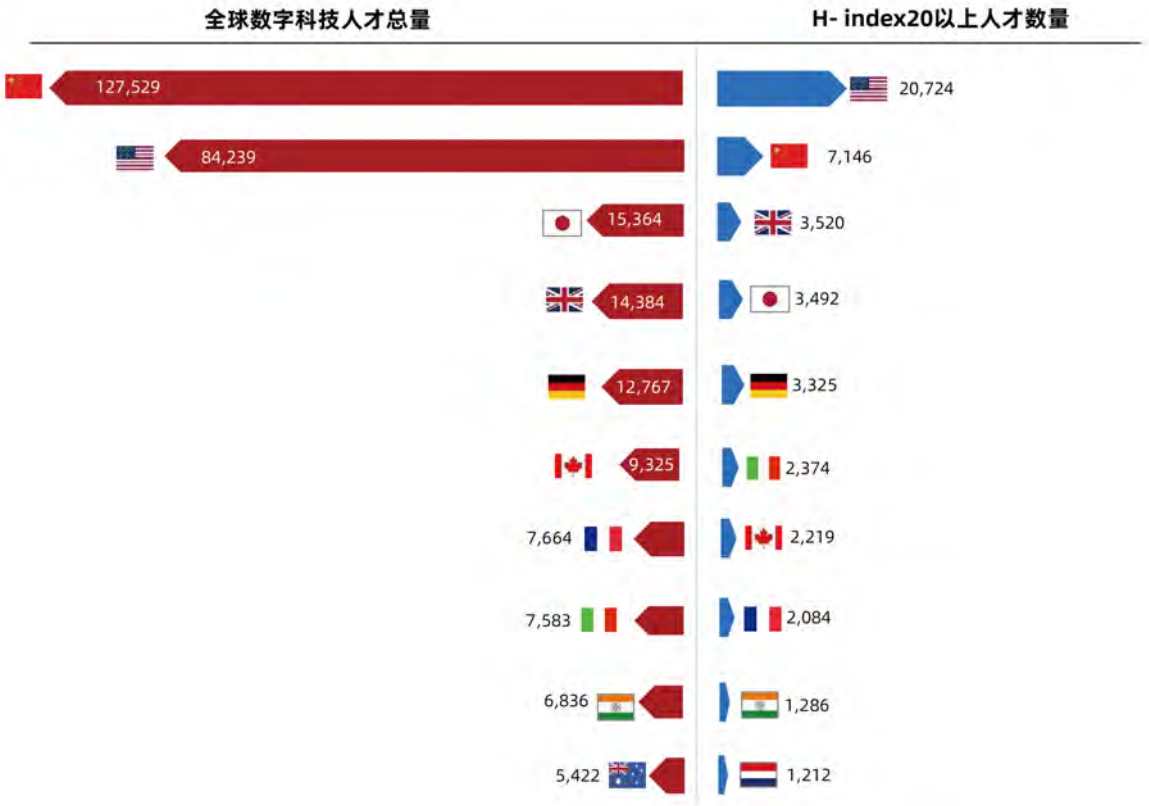
阿里研究院资深技术专家，计算机博士。阿里研究院资深技术专家，阿里研究院未来技术中心负责人，中国中文信息学会理事，中国计算机学会人工智能专委会常务委员。关注下一代互联网技术与产业发展。

报告要点

一、中国人才基数大，但高层次人不足，且严重落后于美国

全球数字科技人才总量为 77.5 万人，其中中国有 12.8 万人，位居第一，占全球总量的 17%，是排名第二的美国的 1.5 倍，是排名第三的日本的 8.3 倍。但中国数字科技高层次人才（即 H-index ≥ 20，下同）只有 0.7 万人，仅占全球总量的 9%，仅为位居第一的美国（2.1 万，占全球 25%）的 35%。这说明我国数字科技人才基数很大，但高层次人才储备不足，且远落后于美国。

图 1 全球数字科技人才数量前 10 强国家



数据来源：AMiner 科技情报平台

二、中国高层次人才集中在高校，而美国高科技公司人才储备不逊色于顶尖高校

从全球数字科技人才机构分布看，中国科学院以 4,722 人的数量名列第一名，以较大优势领先于排名第二的美国加州大学（2,623 人）；进入全球前 10 强的中国机构还有两家，即中国科学院大学（1,462 人，第六名）、清华大学（1,305 人，第八名）。

表 1 全球数字科技人才数量前 10 强机构

全球数字科技人才前 10 强机构				全球 H-index≥20 数字科技人才前 10 强机构			
排名	机构		人才储备（人）	排名	机构		人才储备（人）
1	中国科学院	[中]中国科学院	4,722	1	UNIVERSITY OF CALIFORNIA	[美]加州大学	492
2	UNIVERSITY OF CALIFORNIA	[美]加州大学	2,623	2	Google	[美]谷歌	196
3	Udice Université de Recherche Françaises	[法]UDICE 法国研究型大学联盟	2,534	3	Stanford University	[美]斯坦福大学	178
4	CNRS CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	[法]法国国家科学研究中心	2,298	4	Microsoft	[美]微软	163
5	UNIVERSITY OF LONDON	[英]伦敦大学	1,508	5	中国科学院	[中]中国科学院	144
6	中国科学院大学	[中]中国科学院大学	1,462	6	北京大学 PEKING UNIVERSITY	[中]北京大学	135
7	HARVARD UNIVERSITY	[美]哈佛大学	1,454	7	Carnegie Mellon University	[美]卡内基梅隆大学	133
8	清华大学 Tsinghua University	[中]清华大学	1,305	8	UNIVERSITY of WASHINGTON	[美]华盛顿大学	131

9	TEXAS STATE UNIVERSITY	[美]德克萨斯大学	1,209	**9**	清华大学 Tsinghua University	[中]清华大学	129
10		[俄]俄罗斯科学院	1,123	**10**	上海交通大学	[中]上海交通大学	119

数据来源：AMiner 科技情报平台

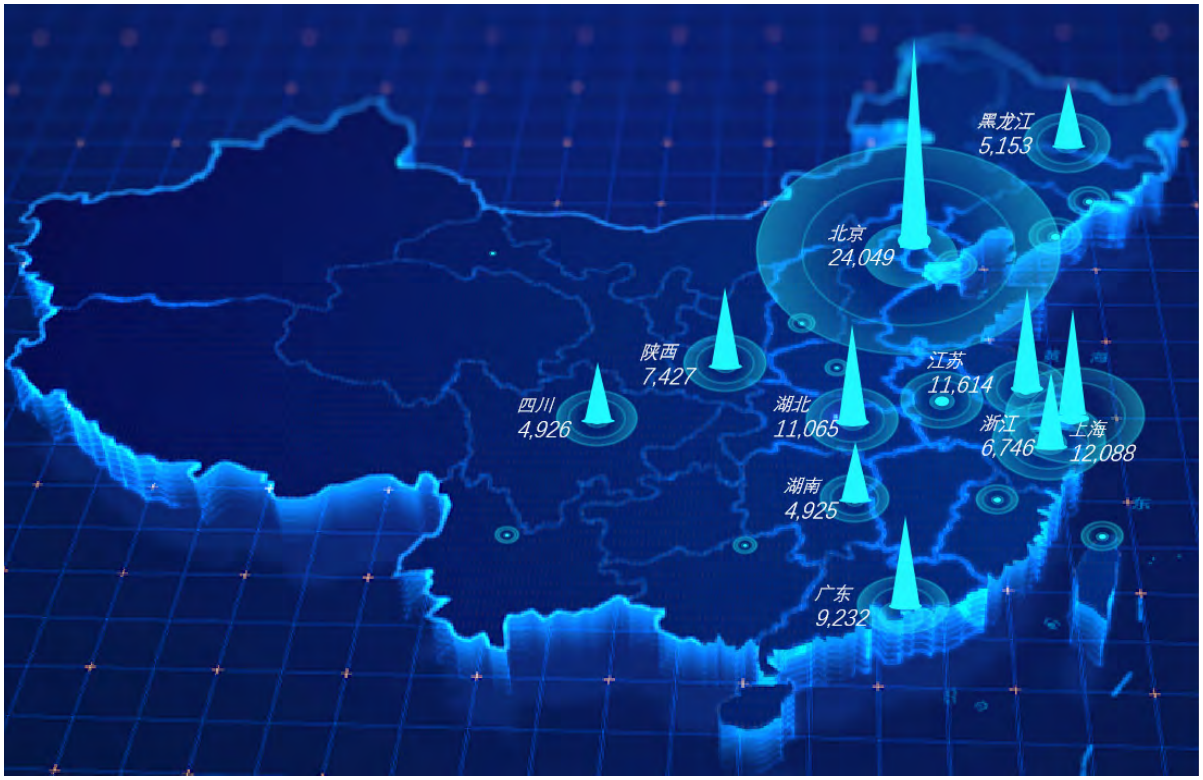
在全球高层次人才前 10 强机构中，美中各占 6 家和 4 家。第一名被美国加州大学摘取，其以 492 人的高层次人才储备大幅度领先于全球其他机构。4 家中国机构依次为中国科学院（144 人，第 5 名）、北京大学（135 人，第 6 名）、清华大学（129 人，第 9 名）、上海交通大学（119 人，第 10 名）。分析发现，中国科学院是中国数字科技人才储备结构的缩影：虽然人才基数大幅领先，但高层次人才明显落后（占其人才总数的 3%，落居全球第五）。

美国相对于中国的另一个优势是，身处科研一线的高科技企业的高层次人才储备实力强于中国，该国的谷歌、微软公司的高层次人才数量排名分别位居全球第二名和第四名，该优势将有力地推动美国数字技术持续创新发展，并引领全球数字技术应用新趋势。

三、北京和上海是中国两大人才高地，前者人才储备为后者 2 倍

从国内数字科技人才分布看，北京人才总数优势明显，共 2.4 万人，占全国近 2 成，是排名第二的上海的 2 倍；北京的高层次人才数量为 1,076 人，占全国 1.5 成，仍然是排名第二的上海的 2 倍。另外，东部地区的广东、江苏、浙江、香港、黑龙江，以及高校资源较丰富的湖北、陕西、四川、湖南等内陆省份也是我国数字科技人才分布较多的地区。

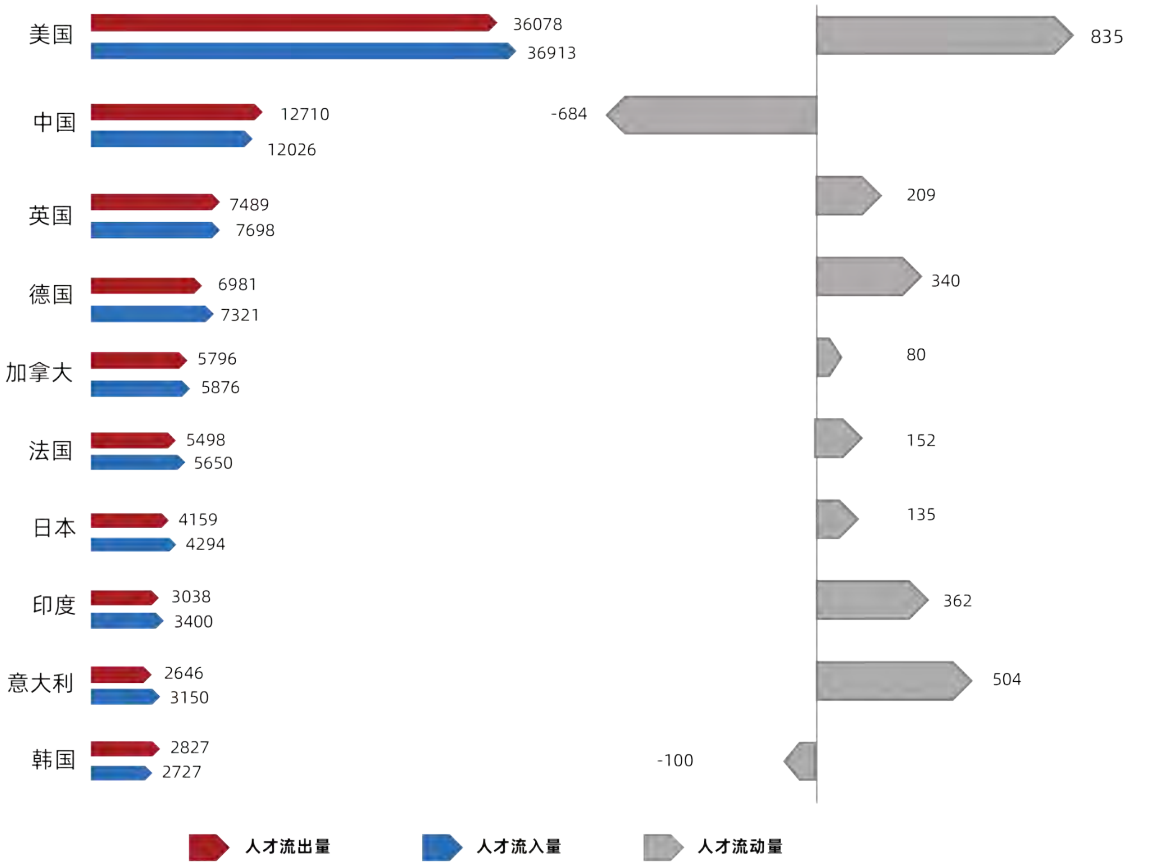
图 2 中国数字科技人才高地（Top 10）



四、中美分别为全球最大的人才净流失国和净流入国

美国是数字科技人才流动最活跃的国家，也是人才净流入最大的国家。过去 10 年，其人才净流入 835 人，位居全球第一；其次是意大利，净流入 504 人。中国则为全球人才净流失第一大国，流失人数达 684 人；其次是韩国，净流失 100 人。由此可见，我国数字科技人才流失严重，在人才引进以及用好、留住人才等方面的工作显得尤为迫切。

图 3 2012-2021 年全球主要国家数字科技领域人才流动对比

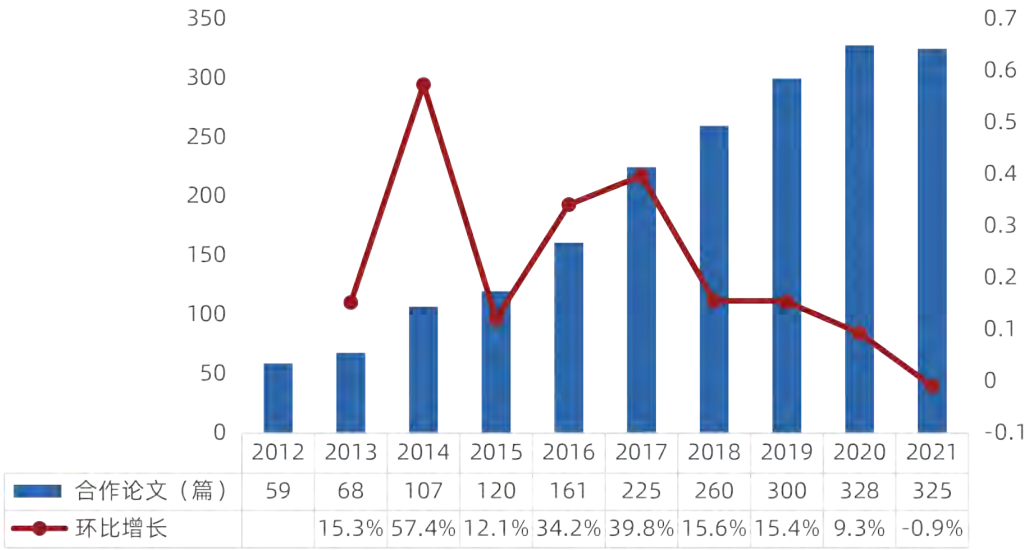


数据来源：AMiner 科技情报平台

五、中美学者合作为全球最热络的双边合作，但从 2018 年开始出现降温

2012-2021 年，在数字科技领域，中美学者共合作产出被引量 Top 1% 的论文 1,954 篇，为全球科研成果产出最丰硕的双边合作（第二为美英学者合作，共产出 1,528 篇）。但从 2018 年开始，中美两国学者合作论文增长率一直处于下降态势，在 2021 年更是出现负增长（-0.9%）。可见，自从 2018 年美国推出所谓“中国行动计划”后，中美在数字科技最卓越研究层次的合作受到严重冲击。

图 4 2012-2021 年数字科技领域中美学者合作关系走势



	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
合作论文（篇）	59	68	107	120	161	225	260	300	328	325
环比增长		15.3%	57.4%	12.1%	34.2%	39.8%	15.6%	15.4%	9.3%	-0.9%

说明：图中合作论文为中美两国学者合作产出被引量 Top1% 的“顶尖论文”数量

六、全球数字科技顶尖科研团队基本被美国机构包揽

全球数字科技顶尖科研团队（Top 10）基本来自美国顶尖高校（如斯坦福大学、卡内基梅隆大学、加利福尼亚大学、麻省理工学院、哥伦比亚大学、康奈尔大学等）、科技巨头（如谷歌、微软、IBM 等）下设的科研机构（如下表所示）。

表 2 全球数字科技顶尖科研团队（Top 10）

排名	顶尖团队		重点研究方向
1		[美]谷歌 X 实验室	人工智能、人机交互
2		[美]斯坦福大学自然语言处理组	自然语言处理
3		[美]卡内基梅隆大学机器学习系	机器学习
4		[美]加利福尼亚大学伯克利分校人工智能研究实验室(BAIR)	人工智能
5		[中]微软亚洲研究院	人工智能、图形学、分布式系统
6		[美]MIT 计算科学与人工智能实验室（CSAIL）	人工智能
7		[美]IBM Thomas J. Watson 研究中心	基础科学、计算机科学
8		[美]哥伦比亚大学数据科学研究所	数据科学
9		[美]康奈尔大学计算科学系	计算机科学
10		[美]美国贝尔实验室	网络、通信

七、我国数字科技人才工作任重道远

首先，高层次人才不足，需加大培养力度。要注重复合培养模式，重点培养有潜力的青年人才；鼓励对外学术交流。其次，人才流失严重，需采取措施引留人才。吸引海外高层次人才特别顶尖华人青年人才是回国发展，用好并留住高层次人才。再次，企业人才储备薄弱，需鼓励企业引育人才。身处科研一线的企业是科技创新的重要主体，有必要鼓励企业加强内部创新环境建设，并在重大专项、重点研发计划论证和实施过程中，支持企业承担政府科研项目，为企业引育人才和提高基础科研能力注入动力。

前言

近年来，数字经济已经成为推动全球经济发展的新动力，各国纷纷出台战略规划，依托数字化技术实现新一轮产业变革，进而抢占全球价值链的制高点。中国政府也做出了重大战略部署，相继出台了多项数字科技发展引导政策，全面推进数字科技发展，并明确提出数字中国战略，构筑国家竞争新优势，加快建成社会主义现代化强国。

2022 年 11 月，习近平总书记在世界互联网大会乌镇峰会的贺信中指出：“当今时代，数字科技作为世界科技革命和产业变革的先导力量，日益融入经济社会发展各领域全过程，深刻改变着生产方式、生活方式和社会治理方式。”国家十四五规划纲要专门设置“第五篇 加快数字化发展 建设数字中国”章节，对加快建设数字经济、数字社会、数字政府，营造良好数字生态做出明确部署。2021 年 12 月，中央网络安全和信息化委员会日前印发《“十四五”国家信息化规划》，提出要建立健全规范有序的数字化发展治理体系，为推动开放、健康、安全的数字生态，加快数字中国建设进程提供方向。

作为技术研发的副产品，学术论文和专利是科研成果的重要结果之一，可以作为研发实力的衡量指标。为了更好地了解全球数字科技的发展情况，我们基于 AMiner[1] 科技情报平台的数据，利用文献计量方法，通过持续跟踪人才发展动态，揭示数字科技的热点前沿领域，对全球各国数字科技发展状况做了全面比较分析。

[1]J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su. Arnetminer: Extraction and mining of academic social networks. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'08), pages 990-998, 2008.

目录

13/ 第一章 全球数字科技人才总体情况

14/ 1. 全球数字科技人才前 10 强国家

15/ 2. 全球数字科技人才前 10 强机构

17/ 第二章 中国数字科技人才分布

18/ 1. 中国数字科技人才前 10 强省市

19/ 2. 中国数字科技人才前 10 强机构

21/ 第三章 数字科技人才流动与合作

22/ 1. 全球数字科技人才流动情况

23/ 2. 中国数字科技人才流动情况

24/ 3. 全球数字科技人才合作情况

27/ 第四章 全球数字科技顶尖科研团队

28/ 1. 谷歌 X 实验室

28/ 2. 斯坦福大学自然语言处理组

29/ 3. 卡内基梅隆大学机器学习系

29/ 4. 加利福尼亚大学伯克利分校人工智能研究实验室 (BAIR)

30/ 5. 微软亚洲研究院

30/ 6. MIT 计算科学与人工智能实验室（CSAIL）

31/ 7. IBM Thomas J. Watson 研究中心

31/ 8. 哥伦比亚大学数据科学研究所

32/ 9. 康奈尔大学计算科学系

32/ 10. 美国贝尔实验室

33/ 第五章 我国数字科技人才问题与建议

34/ 1. 高层次人才不足，需加大人才培养力度

35/ 2. 人才流失严重，需采取措施引留人才

36/ 3. 企业人才储备薄弱，需鼓励企业引育人才

37/ 附件

38/ 1. 研究方法

39/ 2. 数字科技八大领域主题词知识图谱

Chapter 1

全球数字科技人才总体情况

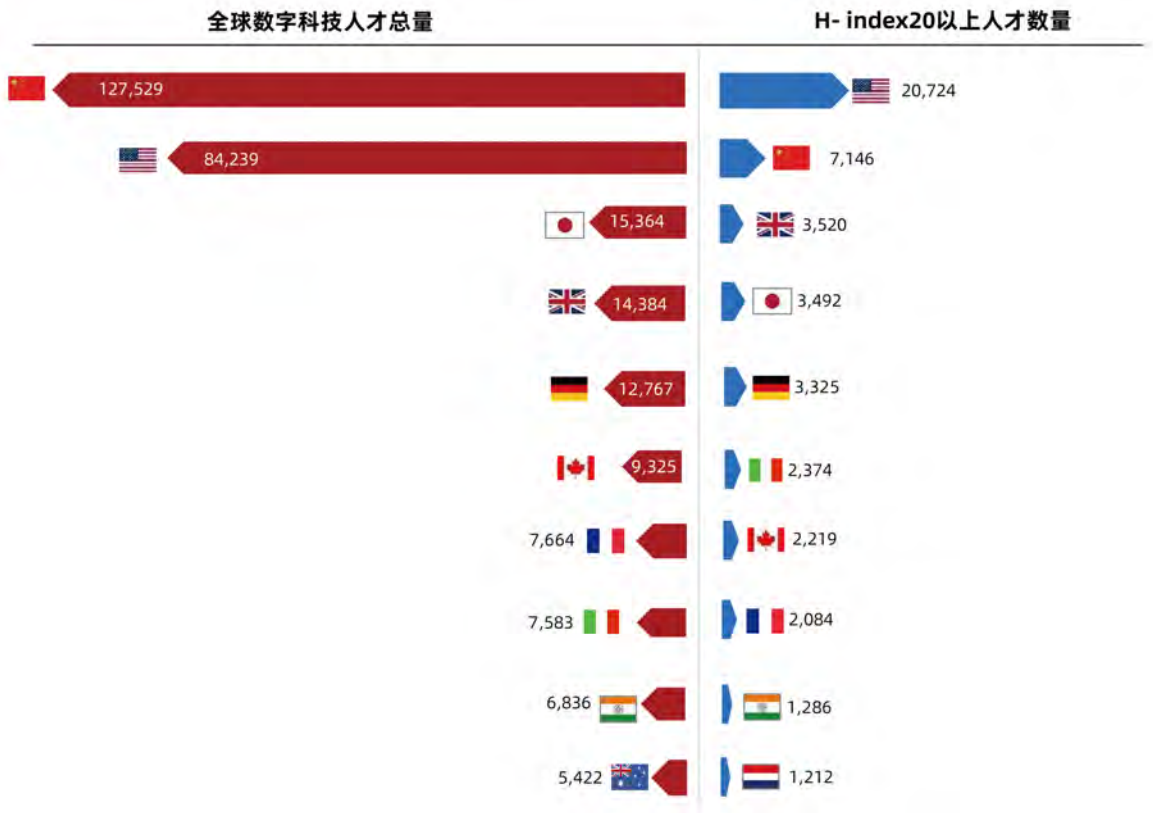
基于 AMiner 科技情报系统，共检索到全球数字科技领域人才 **774,710** 人，其中 H-index[2] 在 20 以上的人才数量为 **83,012** 人。本章节从全球数字领域科技人才的国家和机构分布两个方面进行分析，并总结其特征。

[2]H-index 是一个混合量化指标，目的是量化科研人员作为独立个体的研究成果。一名科研人员 H-index 是指其发表的 Np 篇论文中有 H 篇每篇至少被引 H 次、而其余 Np-H 篇论文每篇被引均小于或等于 H 次。H-index 能够比较准确地反映一个人的学术成就。其数值越高，则表明学者论文的影响力越大，美国学者赫希提出副高职称的 H-index 一般在 20 以上。

1 全球数字科技人才前 10 强国家

全球数字科技人才数量排名前 10 国家如下图所示。其中人数最多的国家是中国，共有 127,529 位，占全球人才总量的 17%；第二名是美国，共有 84,239 位，占全球 11%；第三名是日本，人才数量为 15,364 人，占全球 2%。可见，中国数字科技人才储备优势明显，是排名第二的美国的 1.5 倍，排名第三的日本的 8.3 倍。其他排名前 10 的国家依次为英国、德国、加拿大、法国、意大利、印度和澳大利亚。

图 5 全球数字科技人才数量前 10 强国家



数据来源：AMiner 科技情报平台

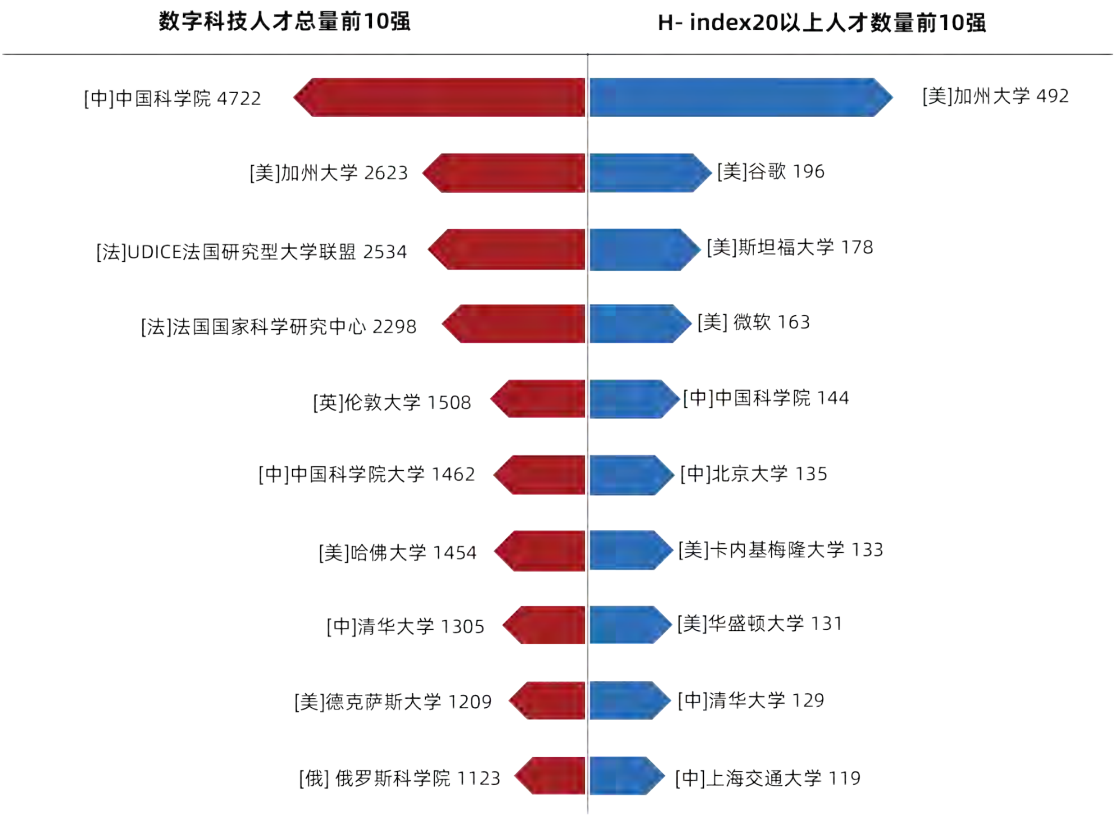
从 H-index 20 以上的人才国别分布看，美国排名第一，共 20,724 人，占全球总量的 25%；中国名列第二，共有 7,146 人，占全球 9%；第三名是英国，共有 3,520，占全球 4%。由此可见，虽然中国数字科技领域人才总量是美国的 1.5 倍，但美国高层次以上人才数量是中国的 2.9 倍。

2 全球数字科技人才前 10 强机构

在全球数字科技人才数量前 10 强机构中，中国科学院以 4,722 人位列第一，加州大学以 2,623 人位列第二，UDICE 法国研究型大学联盟协会以 2,534 人位列第三。其他前 10 强机构依次为法国国家科学研究中心、伦敦大学、中国科学院大学、哈佛大学、清华大学、德克萨斯大学和俄罗斯科学院。

在 H-index 20 以上的人才数量前 10 强机构中，美国和中国各占 6 家和 4 家。中国入榜机构依次是中国科学院（144 人，第 5 名）、北京大学（135 人，第 6 名）、清华大学（129 人，第 9 名）、上海交通大学（119 人，第 10 名）。值得注意的是，1）虽然中国科学院人才总数高居全球第一，但 H-index 20 以上人才（144 人）数量排名下降为全球第五，H-index 20 以上的人才数量仅占该机构人才总数的 3%；2）虽然加州大学人才总数全球排名第二，但 H-index 20 以上人才（429 人）数量以较大优势高居第一，其 H-index 20 以上人才占比为 19%；3）中国进入前 10 强的机构均为高校和科研院所，而美国有谷歌、微软两家高科技公司上榜。

图 6 全球数字科技人才数量前 10 强机构



数据来源：AMiner 科技情报平台

Chapter 2

中国数字科技人才分布

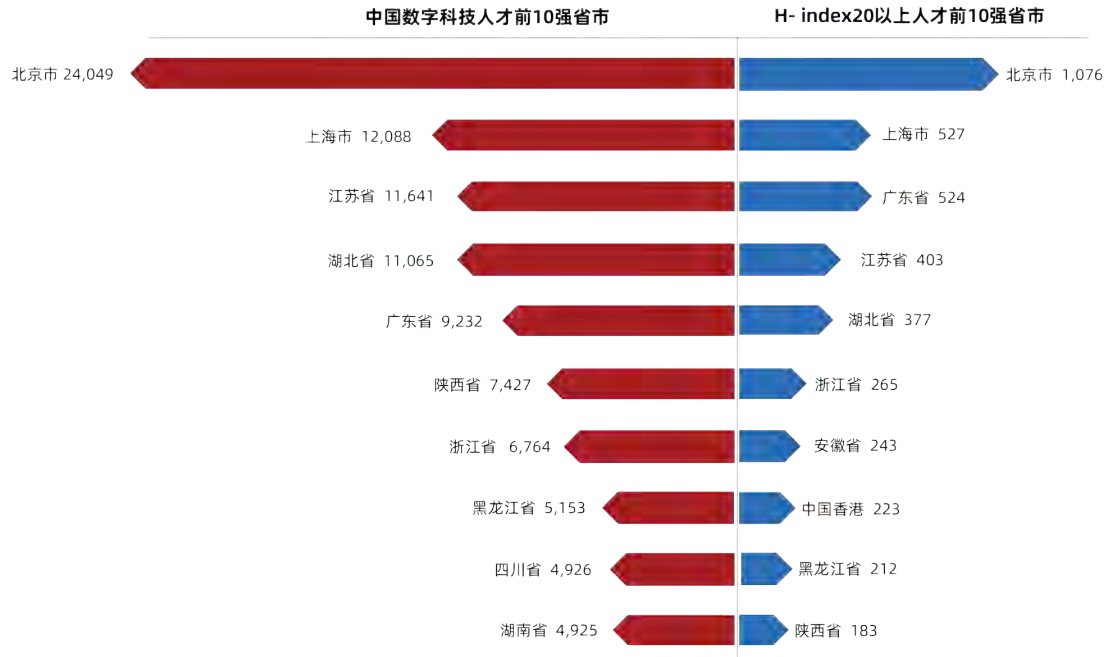
本章节以中国数字科技领域科技人才为研究对象，分别从人才的省市和机构分布两个方面进行分析。

1 中国数字科技人才前 10 强省市

根据 AMiner 科技情报平台显示，北京是全国聚集数字科技人才最多的省市，共有 24,049 人，占全国总量的 19%，是位列第二的上海（12,088 人）的 2 倍。江苏和湖北的人才资源也较多，分别为 11,641 人和 11,065 人。

在 H-index 20 以上的人才储备方面，北京依然排名第一，共有 1,076 人，占全国总量的 15%，仍然是排名第二的上海（527 人）的 2 倍；广东排名第三，共有 524 人。总体看，中国数字科技人才主要分布在经济发达的东部沿海省份，以及重点高校资源较丰富的湖北、安徽和陕西等内地省份。

图 7 中国数字科技人才前 10 强省市



数据来源：AMiner 科技情报平台

2 中国数字科技人才前 10 强机构

从中国数字科技人才机构分布看，中国科学院拥绝对领先优势，共有 4,722 人；中国科学院大学排在第二位，共计 1,462 人；清华大学则以 1,305 名的人才总数紧随其后排在第三位。

在 H-index 20 以上的人才储备方面，中科院、北京大学和清华大学依次以 144 人、135 人和 129 人名列前三甲；其他跻身国内前 10 强的机构依次为上海交通大学、浙江大学、南京大学、哈尔滨工业大学、复旦大学、香港科技大学和中国人民大学。可见除中国科学院和香港科技大学外，上榜的内地高校均为 211 和 985 院校。

表 3 中国数字科技人才数量前 10 强机构

中国数字科技人才前 10 强机构			中国 H-index≥20 数字科技人才前 10 强机构		
排名	机构	人才储备（人）	排名	机构	人才储备（人）
1	中国科学院	4,722	1	中国科学院	144
2	中国科学院大学	1,462	2	北京大学	135
3	清华大学	1,305	3	清华大学	129
4	上海交通大学	1,116	4	上海交通大学	119
5	浙江大学	1,103	5	浙江大学	103
6	北京大学	1,028	6	南京大学	95

7		中山大学	845	7		哈尔滨工业大学	94
8		中国科学技术大学	842	8		复旦大学	66
9		华中科技大学	828	9		香港科技大学	49
10		复旦大学	763	10		中国人民大学	45

Chapter 3

数字科技人才流动与合作

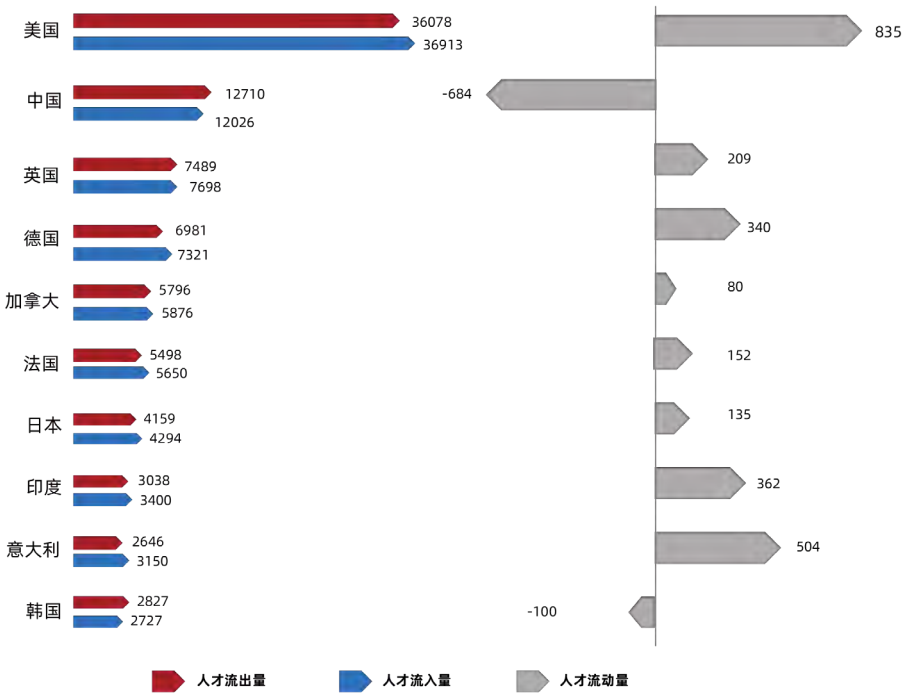
在全球化的背景下，人才合理流动在一定条件下可以优化资源配置，提升技术创新体系整体效能。根据 AMiner 科技情报平台统计，2012-2021 年间，全球数字科技领域人才流动量为 **24.4 万**人次。

1 全球数字科技人才流动情况

从各国数字科技人才流动情况看，美国是人才流动最活跃的国家，流出和流入总量达到 72,991 人次；中国人才流动总量为 24,736 人次，位居第二；英国、德国、加拿大、法国的数字科技才流动也较为活跃，流动总量达 1 万人次以上。

从各国数字科技人才净流入情况看，美国是全球最大的净流入国家，共净流入 835 人；其次是意大利，共 504 人；第三是印度，362 人；另外，德国、英国、法国、日本、加拿大等发达国家均为正向净流入国家。从净流失情况看，中国为全球最大的净流失国家，达 684 人次；其次是韩国，净流失 100 人。

图 8 2012-2021 年全球主要国家数字科技领域人才流动对比



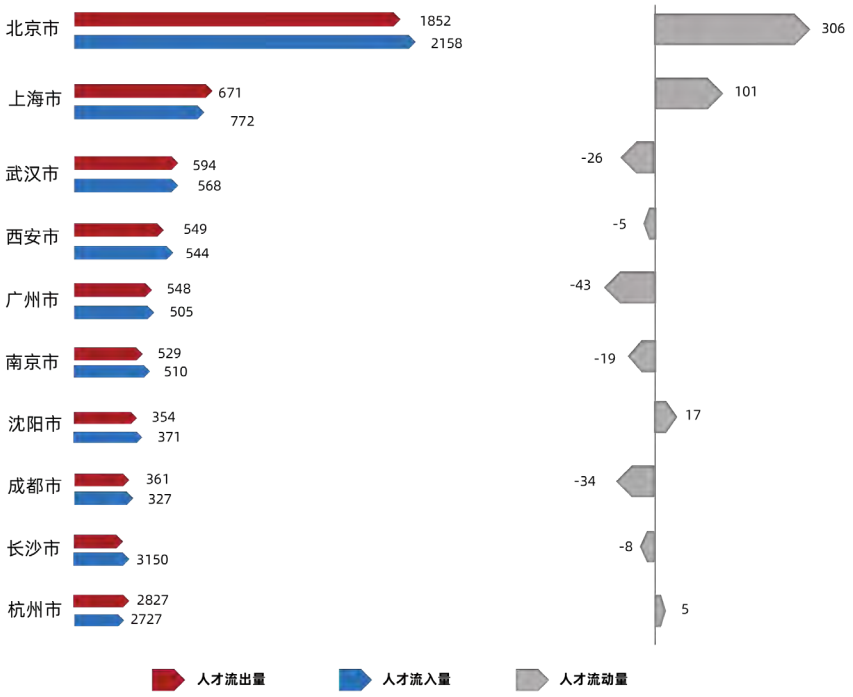
数据来源：AMiner 科技情报平台

2 中国数字科技人才流动情况

从国内数字科技人才流动情况看，北京数字科技人才流动最活跃，流出和流入总量达到 4,010 人次；上海第二，人才流动总量为 1,443 人次；武汉第三，流动总量为 1,162 人次。

从国内各主要城市数字科技人才净流入看，北京最多，共 306 人；其次是上海，共 101 人；另外，沈阳（17 人）和杭州（5 人）也是正向净流入城市。从人才净流失情况看，广州净流失最多，共 43 人；其次是成都，共 34 人；第三是武汉，共 26 人；另外，南京（19 人）、长沙（8 人）、西安（5 人）也出现一定程度的流失现象。可见，我国数字科技人才呈现由内陆城市向京沪一线城市流动的态势。

图 9 2012-2021 年中国主要城市数字科技领域人才流动对比

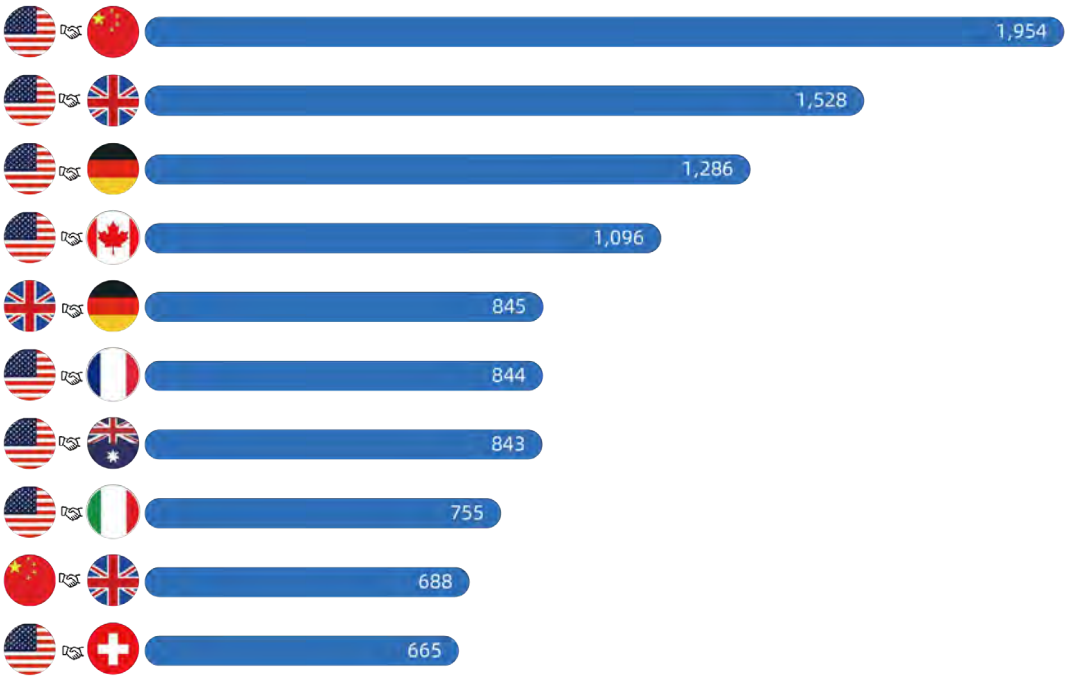


数据来源：AMiner 科技情报平台

3 全球数字科技人才合作情况

论文合作是科研合作最主要的合作形式之一。从各国被引量 Top1% 合作论文数量来看，中美两国的合作的论文最多，共为 1,954 篇；第二名为英美两国，合作数量为 1,528 篇；第三名为美德两国，合作论文数量为 1,286 篇。其他排名靠前的合作国家还有美国和加拿大（1,096 篇）、英国和德国（845 篇）、美国和法国（844 篇）、美国和澳大利亚（843 篇）、美国和意大利（755 篇）、中国和英国（688 篇）、美国和瑞士（665 篇）。整体而言，美国与各国都保持着紧密的合作，在全球合作最热络的前十强双边国家中，有 8 组双边关系由美国与其他 8 个国家组成。

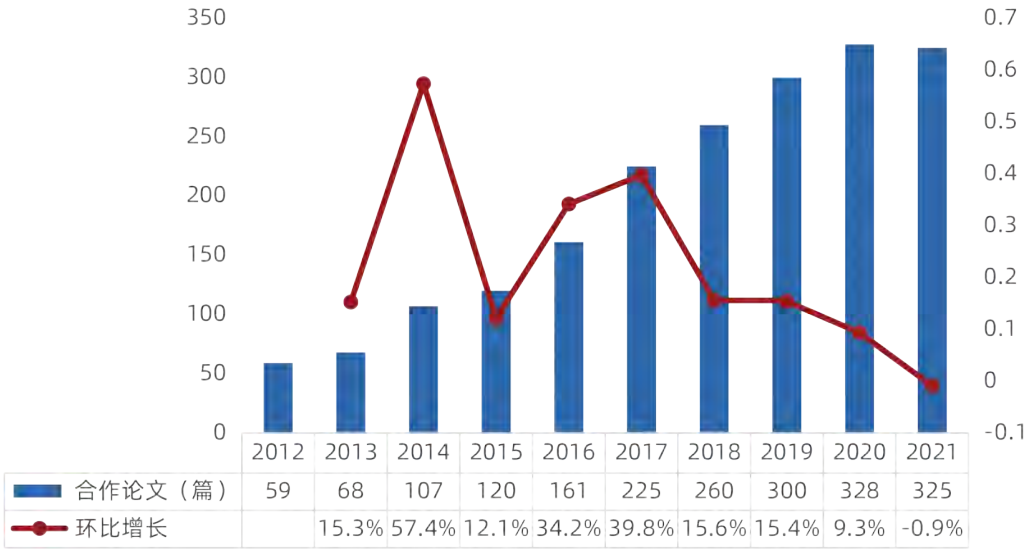
图 10 2012-2021 年全球数字科技人才国际合作关系热络度 Top 10



说明：图中数据为两国学者合作产出被引量 Top1% 的“顶尖论文”数量
数据来源：AMiner 科技情报平台

从中美学者合作的 Top 1% 论文数量的逐年分布来看，从 2012 年开始，中美两国学者合作论文数量一直处于增长态势，到 2020 年达到顶峰（328 篇），2021 年出现小幅下降现象。实际上，两国学者合作论文数量增长率从 2018 年开始就一直处于下降态势，在 2021 年更是出现负增长（-0.9%）。可见，中美两国学者在数字科技最卓越研究层次的合作热络度有所降温，究其原因，这与美国国内对中美在前沿领域科研合作和交流奉行限制性政策，以及全球新冠疫情大流行有关。

图 11 2012-2021 年数字科技领域中美学者合作关系走势



	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
合作论文（篇）	59	68	107	120	161	225	260	300	328	325
环比增长		15.3%	57.4%	12.1%	34.2%	39.8%	15.6%	15.4%	9.3%	-0.9%

说明：图中合作论文为中美两国学者合作产出被引量 Top1% 的“顶尖论文”数量

数据来源：AMiner 科技情报平台


Business Chart - Visual
Business Chart
Interactive User
1,505
18,321
Realtime Dashboard
Space Usage (750 MB)
Target

Chapter 4

全球数字科技顶尖科研团队

坚持全球视野和开放创新，掌握国内顶尖团队及其学者最新概况是攻克行业关键技术或前瞻性技术的重要路径。本篇通过大数据分析和挖掘技术，对数字科技领域机构进行挖掘分析，筛选其中“h5 指数”[3] 排名靠前的十大机构的创新研发团队进行介绍。

[3]h5 指数（h5-index）指是收录在 AMiner 学术系统中的期刊和会议在最近 5 年的论文数量及各论文被引用的次数，可以体现一个机构的综合实力，是评价一个机构影响力的重要参考。例如，如果某机构在过去 5 年所发表的论文中，至少有 h 篇论文分别被引用了至少 h 次，那么这个机构的 h5 指数就是 h。机构排名详情见：https://www.aminer.cn/ranks/org。

1. 谷歌 X 实验室

谷歌公司（Google Inc.）成立于 1998 年，由 Larry Page 和 Sergey Mikhaylovich Brin 共同创建，被公认为全球最大的搜索引擎公司。谷歌旗下有数个研究实验室，其中包括研究方向有些神秘的 X 实验室、研究深度学习的 Google Brain、研究人工智能的 Google AI、研究机器人的 Robotics at Google 以及研究增强现实和智能织物的先进技术和项目集团 (Advanced Technologies and Projects)，自动驾驶汽车 Waymo、互联网气球 Loon、送货无人机 Wing 等研究成果在不同方向上推动了数字科技领域的进步。

重要成员：Astro Teller (人工智能，信号增强算法)、Sebastian Thrun（机器人，人工智能）、Johnny Chung Lee（人机交互）。

机构网址：https://x.company

2. 斯坦福大学自然语言处理组

斯坦福大学自然语言处理组是全球最重要和最突出的实验室之一，由斯坦福语言学系和计算机系共同组成，同时也是斯坦佛大学人工智能实验室的重要组成部分。改组主要研究计算机处理和理解人类语言的算法，工作范围从计算语言学的基本研究到语言处理的关键应用技术均有涉猎，涵盖句子理解、自动问答、机器翻译、语法解析和标签、情绪分析和模型的文本和视觉场景等。该小组的一个显著特征是将复杂和深入的语言建模和数据分析与 NLP 的创新概率、机器学习和深度学习方法有效地结合在一起。

重要成员: Chris Manning (自然语言处理、人工智能、机器学习、信息检索); Dan Jurafsky(自然语言处理)、Percy Liang（计算机科学、人工智能、自然语言处理、机器学习）、Christopher Potts (人工智能、计算机科学、自然语言处理）、Tatsunori Hashimoto (自然语言处理、机器学习）、Monica S. Lam（自然语言处理、深度学习）、Diyi Yang（自然语言处理、人工智能）。

机构网址：https://nlp.stanford.edu

3. 卡内基梅隆大学机器学习系

卡内基梅隆大学机器学习系成立于 2006 年春季，是世界上第一个机器学习系。它是从 1997 年创建的自动学习和发现中心 (CALD) 组织演变而来的。2017 年 6 月，CMU 启动人工智能计划（CMU AI），整合了全校人工智能的研究资源，开展跨学科的人工智能协作，组建世界上规模最大、经验最丰富的 AI 研究团队之一，同时促进本校人工智能方向的人才培养，除覆盖计算机学科以外，还涉及生物、环境工程、哲学、艺术、公共政策等多个领域。

重要成员：Ruslan Salakhutdinov (机器学习)、Maria-Florina Balcan（人工智能、机器学习）、Yuanzhi Li（深度学习）、Henry Chai（机器学习、计算机科学）、Christos Faloutsos（机器学习）、Ziv Bar-Joseph（机器学习）。

机构网址：https://www.ml.cmu.edu

4. 加利福尼亚大学伯克利分校人工智能研究实验室 (BAIR)

加州大学伯克利分校是全球最负盛名的公立学校，人工智能研究 (Berkeley Artificial Intelligence Research, BAIR) 实验主要研究领域涵盖计算机视觉、机器学习、自然语言处理、规划和机器人等。机器人和智能机器实验室致力于用机器人复制动物的行为，自动化科学和工程实验室从事更广泛的机器人功能的研究，如机器人辅助外科手术和自动化制造。计算机可视化小组研究如何让机器人能“看的见”。

重要成员：Trevor Darrell （人工智能、机器人、计算机视觉）、Pieter Abbeel （人工智能、机器人、机器学习）、Alexei Efros （人工智能、计算机视觉）、Peter L. Bartlett（计算机科学、深度学习、神经网络学习）、John Canny（人工智能、机器人）、Ken Goldberg（机器人、计算机科学）。

机构网址：https://bair.berkeley.edu

5. 微软亚洲研究院

微软亚洲研究院是微软公司在亚太地区设立的基础及应用研究机构，也是微软在美国本土以外规模最大的研究机构。微软亚洲研究院从事自然用户界面、智能多媒体、人工智能、云和边缘计算、大数据与知识挖掘、计算机科学基础等领域的研究，致力于推动整个计算机科学领域的前沿发展，将最新研究成果快速转化到微软的关键产品中，着眼下一代革命性技术的研究和孵化。

重要成员：周礼栋（可靠、可信及可扩展的分布式计算系统）、洪小文（自然交互、语音识别）、郭百宁（计算机图形学、计算机可视化、自然用户界面以及统计学习）、刘铁岩（机器学习、信息检索）、潘天佑（智能芯片、信息系统安全）、田江森（智能芯片）、张冬梅（大数据分析、知识计算、数据可视化以及软件工程）、周明（自然语言处理）。

机构网址：https://www.msra.cn

6. MIT 计算科学与人工智能实验室（CSAIL）

CSAIL 是麻省理工学院最大的实验室，专注于开发基础新技术，开展基础研究以推动计算领域的发展研究领域涉及电气工程，计算机科学，数学，航空航天，脑和认知科学，机械工程，媒体艺术与科学，以及地球，大气和行星科学，卫生科学与技术。实验室开创了计算领域的新研究，改善了人们工作、娱乐和学习的方式。

重要成员：Daniela Rus（人工智能、计算机科学、机器人）、Arvind Mithal（人工智能、计算机科学）、Joel Voldman（计算机科学）、Antonio Torralba（人工智能、计算机科学、计算机视觉、机器学习、计算机视觉）、Hal Abelson（计算机科学、人工智能）、Ted Adelson（人工智能、计算机科学）、Anant Agarwal（人工智能、计算机科学）、Pulkit Agrawal（人工智能、计算机科学）、Mohammad Alizadeh（人工智能、计算机科学）、Saman Amarasinghe（人工智能、计算机科学、机器学习）。

机构网址：https://www.csail.mit.edu

7. IBM Thomas J. Watson 研究中心

IBM Thomas J. Watson 研究中心位于美国纽约，专门从事基础科学研究，并探索与产品有关的技术，科学家在这里工作，一方面推进基础科学，一方面提出对实际应用有益的科学新思想。研究范围包括计算机科学、输入 / 输出技术、生产性研究数学、物理学、激光物理、天文学和基本粒子。

重要成员：Dario Gil （人工智能）、Charles Bennett（量子计算、量子密码学、算法信息论）、Rogerio Feris（人工智能、计算机视觉、机器学习）、Mo Yu(于墨)（自然语言处理、机器学习）、Abhishek Kumar（机器学习、人工智能）。

机构网址：https://research.ibm.com/labs/watson

8. 哥伦比亚大学数据科学研究所

美国哥伦比亚大学数据科学研究所致力于培养下一代的数据科学家，并开发创新性技术以造福于社会。该研究所设立了 7 个研究中心，这些研究中心是数据科学转化研究与教育的引擎，也是具备高度商业化潜能的技术的来源。1）网络安全中心专注于开发能在数据整个生命周期保护数据安全与隐私的能力。2）数据、媒体与社会中心（前新媒体中心）关注数据中的人，重点研究如何使用数据来理解人类行为、数据与数据处理如何塑造人们的工作和生活方式、数据对网络世界和数字化世界中的人的意义。

重要成员：Jeannette Wing（人工智能）、Chris Wiggins（应用数学、系统生物学）、Tian Zheng（机器学习）、Matthias Preindl（电力电子系统）、Alan West（电化学）。

机构网址：https://datascience.columbia.edu

9. 康奈尔大学计算科学系

康奈尔大学计算机科学系成立于 1965 年，是最早成立的此类系之一。计算机科学系在理论计算机科学、编程语言、分布式系统、人工智能、信息检索、科学计算和计算机图形等方面处于全球领先地位。

重要成员：Kilian Weinberger （机器学习、数据挖掘）、Kenneth Borup（机器学习、深度学习、计算机视觉）、Todd Schmid（计算机科学）、Weijia Song（云计算）、Christoph Kreitz（计算机科学）。

机构网址：https://www.cs.cornell.edu

10. 美国贝尔实验室

贝尔实验室原名贝尔电话实验室，始建于 1925 年，总部在美国纽约 (后迁至新泽西州的墨里黑尔)。它是一个在全球享有极高声誉的研究开发机构，主要宗旨是进行通讯科学的研究。贝尔实验室自成立以来，共获重大科研成果 50 多项，如有声电影、晶体管、信息论、激光理论、3K 宇宙背景辐射、可视电话、磁泡器件、光通信、数字计算机等。

重要成员：Peter Vetter （可编程网络系统和安全、移动无线电系统、光学系统）、Thierry Klein （先进传感技术、人工智能、机器人、图像和数据分析、自动化）、Gee Rittenhouse（移动通信、无线网络）、Alfred Aho（量子计算、程式语言、编译器、演算法）。

机构网址：http://www.bell-labs.com/

Chapter 5

我国数字科技人才问题与建议

1 高层次人才不足
需加大人才培养力度

我国数字科技人才基数全球第一，占全球 17%，但高层次数字科技人才仅占全球 9%，而美国占 25%，我国仅为美国的 35%。数字科技为世界关键前沿科技，高层次人才落后局面不利于我国数字科技总体能力达到世界领先水平的战略目标。扭转该不利局面，关键是要创新培养机制，提升人才质量，加大力度持续培养本土高层次人才。1）注重复合培养模式。鼓励高校在原有学科基础上拓宽教育内容，注重数字科技各方向交叉融合的复合专业培养新模式。2）重点培养有潜力的青年人才。把培育国家战略人才力量的政策重心放在青年科技人才上，制定实施基础研究人才专项，长期稳定支持一批在数字科技领域取得突出成绩且具有明显创新潜力的青年人才。3）鼓励对外学术交流。鼓励和引导国内科研人才及其团队与全球顶尖数字科技研究机构开展交流与合作。



2 人才流失严重 需采取措施引留人才

我国是全球数字科技人才流失第一大国，而美国是最大的流入国。目前，全球频频上演“高科技战”、“人才争夺战”，在此的背景下，数字科技人才流失问题势必影响我国未来产业发展安全，甚至国家安全，必须引起高度重视。首先，吸引海外高层次人才回国发展。统筹现有人才计划，加强高层次人才特别是海外顶尖华人青年人才的引进工作，对于一时难以回国或来华工作的高层次人才实行柔性引进，采用学术交流、项目合作、技术咨询等方式，柔性引进高层次人才。其次，用好并留住高层次人才。优化科研环境，以信任为基础使用人才，为各类人才搭建干事创业平台，用好用活高层次人才，赋予人才更大的科研自主权，并完善人才评价体系。



3 企业人才储备薄弱 需鼓励企业引育人才

我国数字科技人才集中分布在高校和科研院所。全球数字科技人才前 10 强机构榜单，乃至国内前 10 强机构榜单均不见中国企业身影，而美国高科技巨头谷歌和微软的高层次人才储备实力高居全球第二和第四名。身处科研一线的企业是科技创新的重要主体。因此，1）有必要鼓励企业加强内部创新环境建设，协助企业引进高层次人才，支持企业与高等院校和科研院所共同培养基础研究人才。2）发挥政府科技计划的导向作用，在重大专项、重点研发计划论证和实施过程中，支持企业承担政府科研项目，为企业引育人才和提高基础科研能力注入动力。



Chapter 6

附 件

1 研究方法

数字科技是基于物理世界和数字世界映射互动的体系提炼的一个新概念，是当今世界创新速度最快、通用性最广、渗透性和引领性最强的领域之一。本报告通过人工智能技术，从海量大规模非结构化的数据中抽取数字科技相关知识，生成结构化的知识数据，并根据关系和属性链接等构建数字科技知识图谱，确保了知识体系的完整性，本次研究范围覆盖与数字科技高度相关的八大方向，依次是人工智能、区块链、大数据、云计算、物联网、量子科技、新一代无线通信、XR技术。

本报告围绕数字科技领域的人才分布，利用大数据分析与挖掘、知识图谱、自然语言处理等技术，对领域论文所属的地区和机构的分布格局进行分析。此次分析的时间范围：2012 年 1 月—2021 年 12 月。关键词：采用知识图谱技术抽取，经顾问组专家指导形成领域关键词表，选取领域相关的关键词进行检索。检索范围：自有的大数据服务平台，覆盖全球近 1.4 亿人次人才，3.2 亿篇学术论文。

本报告为系列报告，研究主题分为科研成果、人才资源、技术趋势三部分内容，本期为人才资源篇，以论文和专利成果作为研究对象，对全球数字科技领域科研成果的布局和发展趋势进行分析，供领域人才和从业者客观把握数字科技最新动态。

2 数字科技八大领域主题词知识图谱

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
人工智能	模式识别	文字识别	OCR			
			汉字识别			
		图像识别	人脸识别			
			虹膜识别			
			指纹识别			
		生物特征识别	人脸识别			
			虹膜识别			
			指纹识别			
		语音识别				
	自然语言处理	自然语言处理基础研究	词法分析			
			句法分析			
			语义分析			
			认知语言学			
			语言表示			
			知识图谱			
		自然语言处理应用研究	文本分析			
			文本聚类			
			信息抽取	命名实体识别		
				实体消歧		
				关系抽取		
				事件抽取		
			自动文摘			
			信息推荐			
			自动问答			

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
			机器翻译			
			社会媒体处理			
	计算机视觉	图像处理				
		特征提取				
		图像分割				
		三维重建				
		图像拼接				
		图像识别				
		图像检索				
		目标检测				
		目标跟踪				
		运动行为分析				
		图像标注				
		视频标注				
		虚拟现实				
		增强现实				
	语音技术	语音合成				
		语音识别				
		声纹识别				
		语音增强				
		语音翻译				
		语音系统				
	机器学习	监督学习				
		半监督学习				
		强化学习				
		集成学习	Boosting			
			Bagging			
			随机森林			
			旋转森林			
		数据降维	主成分分析			
			主成分回归			
			因子分析			
			流形学习			
			线性判别分析			
			局部线性嵌入			
			拉普拉斯特征映射			
		度量学习				
		特征选择	稀疏表示			
			字典学习			

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
			压缩感知			
		计算学习理论	VC 维			
		概率图模型	隐马尔可夫模式			
			马尔可夫随机场			
			条件随机场			
			变分推断			
			马尔可夫链			
			马尔可夫蒙特卡			
			洛采样			
			吉布斯采样			
		人工神经网络	神经元模型			
			反向传播			
			多层感知机器			
			深度学习	深度信念网络		
				自由编码器	欠完备自编码	
					稀疏自编码	
					去噪自编码	
					变分自编码	
					对抗自编码	
				卷积神经网络		
				循环神经网络		
				生成对抗网络		
				深度学习框架	Theano	
					PyTorch	
					MXNet	
					Torch	
					TensorFlow	
					Caffe	
		无监督学习	关联规则学习			
			聚类分析	模糊聚类		
				K-means 算法		
				层次聚类		
			k 近邻法			
	数据挖掘	数据预处理				
		关联规则学习	频繁项集挖掘			
			序列模式挖掘			
			高级模式挖掘			
		数据分析	线性模型	线性回归		
				对数几率回归		
				多项逻辑回归		

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
				Softmax 回归		
			排序			
		聚类分析	模糊聚类			
			K-means 算法			
			层次聚类			
		数据异常检测				
		数据降维	主成分分析			
			生成回归			
			因子分析			
			流形学习			
			线性判别分析			
			局部线性嵌入			
			拉帕拉斯特征映射			
		数据可视化				
		数据仓库				
		信息推荐				
		信息检索				
		Kaggle				
区块链	密码学	Hash 算法	MD5、SHA-1 和 SHA-2			
			数字摘要			
		加密算法	对称加密算法	MD5、SHA-1 和 SHA-2		
			非对称加密算法	RSA		
				EIGamal		
			混合加密机制	椭圆曲线		
		数字签名和证书	数字签名			
			数字证书			
		PKI 体系	OpenSSL			
			EJBCA			
			OpenCA			
		Merkle 树	完整性校验			
			SPV 简化支付验证			
			快速比较大量数据			
			零知识认证			
		同态加密	函数加密			
	以太坊	智能合约	对比传统合约			

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
			交易模型			
		协议设计	以太坊虚拟机			
			账户			
			交易			
			一致性			
		开发实战	以太坊钱包	Geth		
			Solidity 语言			
	Hyperledger	组织结构	BlockChain			
			Explorer			
			Fabric			
			Sawtooth Lake			
			Iroha			
			Cello			
		Fabric	机构设计	会员		
				区块链		
				链码		
			安装部署	集群部署		
				部署智能合约		
			开发实战	Dapp 实战开发		
	比特币	中本聪白皮书	账户/地址			
			交易方式			
			脚本系统			
			区块信息			
			双花问题			
			避免攻击			
		挖矿	矿机、矿工			
			算力			
			共识机制	POW		
				POS		
		共识机制	POW			
			POS			
		闪电网络	RSMC			
			HTLC			
			侧链技术			
		交叉	区块扩容			
			软分叉和硬分叉			
	分布式系统	一致性问题	顺序一致性			
			线性一致性			
			CAP 原理			
			FLP 不可能原理			

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
		分布式系统共识	Paxos 算法			
			Raft 算法			
		拜占庭问题与算法	拜占庭问题			
			BFT 算法			
大数据	基础技术	数据分片式路由				
		数据复制&一致性				
		大数据常用算法与数据结构				
	数据采集	系统日志采集				
		外部数据采集				
		IOT 设备数据采集				
	数据传输	消息队列				
		数据同步				
		数据订阅				
		序列化				
	数据存储	物理存储				
		分布式文件/对象存储系统				
		分布式关系型数据库				
		分析型数据库				
		搜索引擎				
		键值存储数据库				
		图数据库				
		列存储数据库				
		文档数据库				
		时序数据库				
		RDF 数据库				
		多媒体数据库				
		事件存储数据库				
	数据处理	数据计算				
		数据仓库				
		数据湖				
		数据分析				
		数据挖掘				
		平台工具				
	数据应用	数据服务				
		数据可视化				
		数据共享				
		数据预警				

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
	数据治理	数据规范标准				
		元数据管理				
		数据质量				
		主数据管理				
		大数据架构管理				
		数据安全				
		业界框架产品				
云计算	SaaS	Web Server	Nginx	构架原理		
				负载均衡		
				反向代理		
				正向代理		
			Tomcat	架构原理		
				调优方案		
		缓存	本地缓存			
			客户端缓存			
			服务端缓存	Memcached		
				Redis		
		消息队列	RabbitMQ			
			RocketMQ			
			ActiveMQ			
			kafka			
			Redis 消息推送			
		搜索引擎	ES 的分布式架构原理			
			倒排索引			
			集群化部署			
	LaaS	虚拟化技术	安全虚拟化	软件辅助虚拟化	CPU 虚拟化	
					内存虚拟化	
					I/O 虚拟化	
				硬件辅助虚拟化	CPU 虚拟化的硬件支持（VT-x）	
					内存虚拟化的硬件支持（EPT）	
					IO 虚拟化的硬件支持（VT-d）	
					时间虚拟化	
			类虚拟化	CPU 虚拟化		
				内存虚拟化		
				I/O 虚拟化		
				时间与时钟管理		

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
		OpenStack	OpenStack 体系架构	OpenStack 发展历程		
				OpenStack 组件关系		
			计算组件	Nova 体系结构		
				Nova API		
				Rolling Upgrade		
				Shueduler		
				典型工作流程		
			存储组件	Swift		
				Cinder		
				Glance		
				Ceph		
			网络组件	Neutron 体系结构		
				Neutron API		
				网络插件	ML2 Plugin	
					Port Binding	
					Open vSwitch Agent	
					Service Plugin	
				其他	DVR	
					SDN	
					NFV/SRIOV	
					OVS & DPDK	
			安全组件	Keystone 体系结构		
				Keystone 启动流程		
				令牌及证书相关		
				可信计算池		
			计量与监控	Ceilometer		
				Aodh		
				Gnocchi		
				Panko		
			OpenStack 部署与运维	OpenStack 部署	Devstack 一键部署	
					docs.OpenStack.org 分布式部署	
					扩容&升级	
				Openstack 运维	Horzion 控制面	

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
					板	
					维护&故障&调试	
					网络排障&日志监控	
					备份和恢复	
	PaaS	Docker	Docker 基础	Docker 安装		
				Docker 基本操作		
			Docker 核心原理	内部资源隔离	namespace	
					cgroups	
				Docker 架构	daemon	
					client	
					镜像管理	
				libcontainer		
				存储管理		
				数据卷		
				网络管理	Docker 网络基础	
					Docker daemon 原理	
					libcontainer 原理	
					link 原理	
					pipwork 原理	
				容器安全		
			Docker 编排三剑客			
				Compose		
				Machine		
				Swarm		
		Kubernetes	Kubernetes 体系架构	Kubernetes 设计解读		
				Kubernetes 基本组件		
			Kubernetes 组件原理			
			Kubernetes 部署与运维	Kubernetes 部署	Kubernetes 快速部署	
					二进制文件部署	
				Kubernetes 运维	重置&升级	
					Kubernetes 集群管理	Dashboard
						Kuboard
						Prometheus & Grafana

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
						Helm 包管理工具
					Trouble Shooting	
			Kubernetes 开发指南	Kubernetes API 详解		
				Kubernetes AIP 扩展		
				新功能	调度器	
					Pod 的垂直扩缩容	
					GPU 支持	
		Kubeedge	边缘计算场景			
			Kubeedge 安装部署	X86		
				arm32 & arm64		
			Kubeedge 架构及其组件	Kubeedge 架构原理		
				Edge 端组件	EdgeD	
					EventBus	
					MetaManager	
					Edgehub	
					Device Twin	
				Cloud 端组件	Edge Contoller	
					CloudHub	
					Device Controller	
物联网	感知识别层	芯片	系统级芯片			
			传感器芯片			
			基带芯片			
			射频芯片			
		传感器	物理传感器			
			化学传感器			
			生物传感器			
		感知设备	RFID 读写器			
			条码扫描器			
			卫星定位装置			
			语音识别设备			
			视觉成像设备			
	网络传输层	通信模组	蜂窝通信模组	5/6G		
				NB-IoT/eMTC		
			非蜂窝类模组	WiFi 模组		

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
				蓝牙模组		
		通信网络	蜂窝通信网络	5/6G		
				NB-IoT/eMTC		
				SIM 卡		
			非蜂窝通信网络	WiFi		
				蓝牙		
				Zigbee		
				LoRa/Sigfox		
		基础通信设施	基站			
			天线系统			
			服务器			
	平台管理层	物联网平台	连接管理平台			
			设备管理平台			
			应用使能平台			
			业务分析平台			
		系统&软件	操作系统			
			软件开发			
	应用服务层	智能终端	可穿戴设备			
			智能家电			
			消费电子			
			智能仪表			
			无人机			
		系统集成应用服务	车联网			
			智能家居			
			智慧城市			
			智慧农业			
			远程医疗			
量子科技	量子计算	量子因数分解				
		量子搜索				
	量子通信	量子秘钥分发				
		量子隐性传态				
		通信终端	量子安全	量子秘钥管理模块		
				终端加密		
				终端解密		
			终端通信			
			通信模式选择模块			
	量子测量	量子目标识别	量子干涉雷达			
			量子照射雷达			

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
			量子增强雷达			
		量子重力测量	喷泉式原子重力仪			
			自由下落式重力仪			
		量子磁场测量	光泵磁力仪			
			原子 SERF 磁力仪			
			相干不居囚禁磁力仪			
			金刚石色心磁力仪			
		量子定位导航	量子导航系统	六星定位系统		
				三星定位系统		
			量子加速度计	原子干涉加速度计		
				金刚石色心加速度计		
			量子陀螺仪	原子干涉陀螺仪		
				超流体干涉陀螺仪		
				核磁共振陀螺仪		
				SERF 陀螺仪		
				金刚石色心陀螺仪		
		量子时间基准	量子时钟源	冷原子喷泉钟		
				原子/离子光钟		
				相干布囚禁钟		
			量子同步协议	量子保密时间同步协议		
				纠缠消除色散时间同步协议		
				符合测量纠缠光子对时间同步协议		
				HOME 干涉时间同步协议		
新一代无线通信技术	超密集组网	关键技术	多链接技术			
			无线回传技术			
		部署架构	微基站+微基站			
			宏基站+微基站			
	动态自组织网络：					

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
	SON					
	网络功能虚拟化：NFV					
	5G/6G					
	终端设备					
	eMBB 大流量移动宽带业务					
	mMTC 大规模物联网业务					
	uRLLC 无人驾驶					
XR（虚拟现实、混合现实、增强现实）	计算机图形学	数学基础	向量			
			齐次坐标			
			几何变化			
		建模	几何造型			
			曲线/曲面			
		渲染	消隐			
			光栅化			
			光照模型			
		交互技术	二维图形交互技术			
			三维图形交互技术			
	计算机仿真	仿真建模技术	机理建模			
			辨识建模			
			物理建模			
			智能建模			
			多媒体建模			
		仿真支持系统与平台技术	仿真语言与工具			
			网络化仿真平台			
			仿真资源库管理			
			仿真计算机系统			
			仿真可视化系统			
			仿真虚拟器			
		仿真应用技术	仿真应用共性技术	仿真 VV&A 技术		
				仿真实验技术		
				仿真实验结果分析与评估技术		
			应用领域仿真技术	自然科学与工程信息技术		
				社会科学仿真技术		

一级主题词	二级主题词	三级主题词	四级主题词	五级主题词	六级主题词	七级主题词
				管理科学仿真技术		
				生命科学仿真技术		
				军事科学仿真技术		
	人机接口	脑信号记录				
		信号处理				
		模式识别				
		机器学习				
		感知反馈				
		刺激信号处理				
		脑刺激				
		基础神经科学				
	传感技术	光声光谱传感技术				
		分布光纤传感技术				
		法拉第磁光传感技术				
		液态金属传感技术				
		磁阻电流传感技术				
	多媒体技术	视觉	位图图像	动态影像 真实感三维动画		
			动态图像			
			矢量图形			
			文字			
		听觉	语音（人类语言）			
			声响（自然界）			
			音乐			
		触觉	振动			
			运动			