人工智能与新一轮科技革命

王战 郭进 著

上海人民出版社

人工智能与新一轮科技革命

王战 郭进
著

上海人民出版社

图书在版编目(CIP)数据

人工智能与新一轮科技革命/王战,郭进著. —上海:上海人民出版社,2023
ISBN 978-7-208-18519-7

Ⅰ. ①人… Ⅱ. ①王… ②郭… Ⅲ. ①人工智能-关系-新科技革命-研究 Ⅳ. ①G301

中国国家版本馆 CIP 数据核字(2023)第 166465 号

责任编辑 宫兴林
封面设计 水玉银文化

人工智能与新一轮科技革命
王 战 郭 进 著

出　　版 上海人民出版社
(201101 上海市闵行区号景路 159 弄 C 座)
发　　行 上海人民出版社发行中心
印　　刷 上海商务联西印刷有限公司
开　　本 720×1000 1/16
印　　张 16
插　　页 4
字　　数 139,000
版　　次 2023 年 10 月第 1 版
印　　次 2023 年 10 月第 1 次印刷
ISBN 978-7-208-18519-7/C·700
定　　价 72.00 元

前　言

党的二十大报告指出："当前，世界百年未有之大变局加速演进，新一轮科技革命和产业变革深入发展，国际力量对比深刻调整，我国发展面临新的战略机遇。"要牢牢掌握发展主动权，把握好新的战略机遇，必须抢抓新一轮科技革命和产业变革带来的重大机遇，全面提升对科技创新的战略谋划和快速响应能力，从而抢占未来发展的制高点，赢得先机和主动。

上海如何进一步强化全球资源配置功能、科技创新策源功能、高端产业引领功能、开放枢纽门户功能，深化"五个中心"建设的战略重点，"下好先手棋，打好主动仗"。对此，十二届上海市委三次全会强调，要强化科技创新策源功能，激发高质量发展的澎湃动力，加强基础研究，强化关键核心技术攻关和科技成果转移转化，构建具有全球竞争力的开放创新体系。

为帮助社会大众更好地认识新一轮科技革命和产业变革的趋势，以及上海科技创新中心建设的机遇、挑战和方向，我们邀请著名战略家、上海市社会科学界联合会主席、上海

国际经济交流中心理事长王战教授主持，上海国际经济交流中心研究部部长郭进执笔撰写了这本《人工智能与新一轮科技革命》。吴真如博士参与了场景产业案例的编写。

人工智能是新一轮科技革命和产业变革的重要驱动力量，加快发展新一代人工智能是事关我国能否抓住新一轮科技革命和产业变革机遇的战略问题。本书聚焦新一轮科技革命的主攻方向人工智能，从新一轮科技革命与中国式现代化、科技革命的演进历程和最新趋势、人工智能革命的金字塔框架图、人工智能革命的全球竞争，上海科技创新中心建设五个方面，系统剖析了人工智能革命的兴起背景、发展脉络、技术模式、产业体系和场景应用等重要主题。全书内容丰富，深入浅出，具有很强的前瞻性、学理性和可读性。

本书是上海市社会科学界联合会和上海人民出版社共同策划推出的“望道计划”系列通俗理论读物之一。该计划将陆续出版一批呼应国家战略、紧扣时代实践，坚持学术性、注重大众性的学术读物，努力以更宽广的视野、更专业的分析、更有温度的人文关怀，为广大读者更好理解关于推进拓展中国式现代化和建设中华民族现代文明等重大命题的深刻内涵，提供有益参考。

上海市社联党组书记、专职副主席　王为松

2023 年 9 月

目 录

第一章

新一轮科技革命与中国式现代化

进入21世纪以来，全球科技创新进入空前密集活跃的时期，新一轮科技革命和产业变革正在重构全球创新版图和经济结构。[1]以信息、生命、制造、能源、空间、海洋等领域的前沿技术突破为标志，科学和技术之间、技术与产业之间日益呈现交叉融合的趋势，新一代信息技术加速突破应用，生命科学领域孕育新的变革，传统制造业向智能化、服务化、绿色化转型，全球能源体系加速变革，人类生存的空间不断拓展。正如习近平总书记指出：科学技术从来没有像今天这样深刻影响着国家前途命运，从来没有像今天这样深刻影响着人民生活福祉。[2]

新一轮科技革命以人工智能技术的颠覆式创新为先导，开启了机器智能替代人脑智慧的进程。如果说此前的科技

[1] 习近平：《努力成为世界主要科学中心和创新高地》，《求是》2021年第6期，第1—10页。

[2] 习近平：《努力成为世界主要科学中心和创新高地》，《求是》2021年第6期，第1—10页。

革命和产业变革都是人类运动功能的扩展和延伸，那么人工智能作为新一轮科技革命的先导产业，其核心技术就是对人类智慧的模仿、超越和替代，将从顶层设计和底层逻辑两个层面，同时推动人类经济、社会、文化领域的变革，极大地改变人类现有的生产、生活方式，以及知识的形成和传承模式，也难免对维系人类社会的伦理道德带来极大的冲击，甚至决定人类文明的走向。

鉴此，本书将新一轮科技革命称为类脑技术革命，其内容和体系可以用“金字塔框架图”来归纳和描述，我们将在后文详细论述。

一、科技革命决定了国家的兴衰

科技革命决定国家兴衰，党和国家对此有着清醒的认识。早在2013年9月30日，习近平总书记在主持十八届中央政治局第九次集体学习时，就敏锐地指出：“新一轮科技革命和产业变革正在孕育兴起。”此后，总书记又在国际国内多个重大场合发表了一系列重要讲话，深刻地揭示了新科技革命的本质、内涵及其对人类经济社会发展的深远影响，并就我国如何迎接新科技革命进行了部署。其中，在2016年4月国家网络安全和信息化工作座谈会上，习近平总书记进一步细致分析了新科技革命的深刻影响，指出

“从社会发展史看，人类经历了农业革命、工业革命，正在经历信息革命。农业革命增强了人类生存能力，使人类从采食捕猎走向栽种畜养，从野蛮时代走向文明社会。工业革命拓展了人类体力，以机器取代了人力，以大规模工厂化生产取代了个体工场手工生产。而信息革命则增强了人类脑力，带来生产力又一次质的飞跃，对国际政治、经济、文化、社会、生态、军事等领域发展产生了深刻影响”。

正如习近平总书记在2014年两院院士大会发言时指出“一个国家是否强大不能单就经济总量大小而定，一个民族是否强盛也不能单凭人口规模、领土幅员多寡而定。近代史上，我国落后挨打的根子之一就是科技落后”。在2018年的金砖国家工商论坛上，习近平总书记强调：“在新科技带来的新机遇面前，每个国家都有平等发展权利。潮流来了，跟不上就会落后，就会被淘汰。”

当前，我们又一次迎来了新科技革命的历史关口，面对稍纵即逝的历史机遇，如何“下好先手棋，打好主动仗”，事关国家和民族的未来。如果再次错失新科技革命的历史机遇，中国很可能像过去百年一样，被牢牢锁死在产业链的低端，成为新全球体系的边缘国家，甚至可能付出更大代价。因此，必须精心谋划和超前布局，积极抢占新科技革命的发展先机和技术制高点，“从而使我们成为新的

竞赛规则的重要制定者、新的竞赛场地的重要主导者”[1]。机不可失，时不再来，面临百年未有之大变局，必须紧紧抓住科技革命和产业变革的潮流，乘势而上、掌握主动，中国崛起可期，民族复兴有望。

二、科技革命为大国崛起提供了机遇

回顾历史，不难发现科技革命离不开雄厚的物质基础、众多优秀的科技人才、蓬勃向上的民族创新精神、优越的地理环境和巨大的优质市场。科技革命不仅是企业的竞争，技术的竞赛，更是国力的比拼。此前的历次科技革命，以英美为首的西方世界建立了完整的知识产权体系，配合强大的金融网络、司法体系、军事霸权和跨国公司，西方世界基本锁死了后发国家的赶超路径。只有打破西方世界的这种知识产权和产业竞争的壁垒，后发国家才可能赢得新的空间，掌握发展的主动权。然而，西方世界用两百年时间构筑的知识产权壁垒非常强大，已经渗透到每一个细分领域、每一个技术环节，沿着既有的技术轨道，很难跳出旧体系的制约。可以说，没有知识产权，就没有核心利益。

新科技革命的到来，为中国打破西方重重封锁提供了

[1]《牢牢把握新一轮工业革命的机遇——学习习近平总书记关于新工业革命的重要论述》，《红旗文稿》2018 年第 20 期。

机遇。新科技革命涉及科学和技术的深刻变革，“新一轮科技革命和产业变革正在积聚力量，催生大量新产业、新业态、新模式，给全球发展和人类生产生活带来翻天覆地的变化”[1]。新的技术模式、新的技术创新，意味着新的知识体系、新的专利框架，这为中国打破西方限制，提供了历史性的战略机遇。具体而言有以下四个方面：

第一，新科技革命产生大量颠覆性的研究成果，将催生出全新的产业。新产业意味着新机遇，旧势力无法及时在新领域形成专利壁垒、技术优势和市场垄断，为新势力的技术创新、产业创造提供了机遇。比如，中国在信息技术革命中，起步时间比领先国家较晚。此后，虽然采取了技术引进、跟踪模仿等策略，不断缩小差距，但后发赶超的成效始终不尽人意，其中很重要的一条原因就是西方构筑了强大的技术壁垒和产业优势，使得中国向产业链两端升级的努力异常艰难。这样的案例不胜枚举，比如燃油车、高端精密机床等，旧的技术赛道很难给中国式现代化提供广阔的发展空间。作为一个大国，中国需要开拓科技革命的星辰大海，用新技术、新赛道开创美好生活。只有抢占新一轮科技革命的发展先机，获得标志性科研成果和核心专利，并以此构建中国的现代产业体系，中国才能打破西

[1] 习近平：《努力成为世界主要科学中心和创新高地》，《求是》2021 年第 6 期，第 1—10 页。

方世界的封锁和打压，开启新的发展征程。

第二，新科技革命将带来全新的生产方式，大大降低生产成本。新科技革命能够大幅度降低了制造成本特别是人工成本，比如3D打印技术能够精确控制原材料投入，增材制造的模式大大减少了生产浪费。人工智能技术和智能制造系统能够大大缓解了劳动用工压力，比如智能机器人的广泛应用，能够减少高危、高强度环境中的用工比例，在提升效率的同时，也大大减少了人员的成本开支。智能管理系统的普及，能够建立网络化的协同模式，人类在生产组织环节的浪费现象和无效协同将大为减少。

第三，新科技革命对全球产业分工产生巨大影响。新科技革命，尤其是人工智能的革命，在改变制造业成本的同时，也使得发展中国家丧失了低成本的优势。因为物流成本节约、生产成本较低等原因形成的海外设厂模式和全球产业分工体系，可能就此被扁平化的智能制造供应链取代，制造业回流发达国家的障碍将不复存在，发展中国家实现工业化的战略机遇可能就此长期丧失。同时，新科技革命的领先国家也能够在新的分工体系中，凭借技术优势构建新的发展优势。

第四，新科技革命改变了全球市场经济模式。数字制造技术包括了对新产品进行构思的能力、利用设计测试器和模拟器进行测试的能力、利用计算机软件对三维对象进

行具体设计的能力，以及利用三维打印成型系统（RP 快速原型）对设计进行制作的能力。这些技术的综合运用使产品研发设计、生产作业、控制管理等整个制造系统发生根本改变。新的生产工艺为制造业小批量、个性化生产创造了前提条件，从而要求生产基地更加靠近消费市场，以便设计人员能够和生产线更密切地合作，同时更能贴近客户并迅速回应其需求。能否把市场多样化的需求快速反映到产品的设计当中来并满足多样化需求，将成为企业核心竞争力之所在。

上述四个方面的改变，不但给中国及其他新兴国家带来了机遇，也对现有的国际秩序、经济体系和竞争模式带来巨大冲击，对目前拥有传统竞争优势的在位企业冲击尤其巨大。能否适应新的技术环境、构建新的产业优势，成为决定传统大国、传统企业命运的大事，这是新兴国家、新兴企业崛起的百年机遇。

三、中国能够抓住新一轮科技革命的机遇

党的十八大以来，我国的科技创新能力和产业实力快速提升，已成为具有重要影响力的科技大国和全球重要的生产制造基地，形成了全球规模最大的研发人员队伍，打造了完整的工业体系和创新链条，具备了跨越式发展的基

础和条件。当前，国家已进入量能积累到提质增效的关键发展阶段，完全具备了抓住新科技革命机遇的可能。

人工智能革命是新科技革命的核心。作为人脑智慧的延伸，人工智能的国际竞争不仅是综合国力的较量，更是技术创新与产业融合的综合比拼。一方面，经过60多年的演进，人工智能的发展进入新阶段，成为国际竞争的新焦点。当今世界，主要发达国家把发展人工智能作为提升国家竞争力、维护国家安全的重大战略，加紧出台规划和政策，围绕核心技术、顶尖人才、标准规范等强化部署，力图在新一轮国际科技竞争中掌握主导权。另一方面，我国在人工智能领域的发展上有其独特优势，如稳定的发展环境、充足的人才储备、丰富的应用场景等。中国在技术创新与产业融合领域具有天然优势。

未来，中国需要以人工智能等突破性技术创新为先导，持续优化创新网络和创新制度建设，深化科技体制机制改革，从创新网络、创新机制、产业模式和激励政策等诸多方面同步发力，积极迎接新科技革命的挑战。

第二章

科技革命的演进历程和最新趋势

本轮科技革命的鲜明特色是以人工智能为代表的类脑智能革命，目前正由弱人工智能发展为强人工智能，未来有可能进化为超级人工智能。本书用“金字塔框架图”来概括人工智能为引领的新科技革命，后续章节将详细介绍其主导的人工智能技术、新基建、新平台和场景应用。本章将首先梳理科技革命的发展历程、趋势特点以及历史影响。

一、新科技革命的时代特征

（一）颠覆性技术的群体突破与融合创新

新一轮的科技革命表现出颠覆性技术突破、多领域融合创新的特征。一是多个领域出现了颠覆性的重大技术突破。人工智能、航天科技、量子通信、生物技术等众多领域都出现了影响世界的颠覆性技术，总体呈现数字化、网

络化、智能化、绿色化的发展趋势。其中，人工智能和5G通信等新一代信息技术已经进入产业化的实用阶段，基因科学、生物医药、干细胞再生等领域也取得突破性进展。新能源、航天科技等领域的技术突破频繁出现，量子通信、工业互联网等前沿领域的创新日趋实用化，呈现群体性突破态势。二是技术创新的交叉融合现象更加明显。科学与技术之间、技术与技术之间边界日益模糊，多领域技术交叉融合创新层出不穷。人工智能被广泛应用于各个领域，极大地改变了科技创新的传统范式，融合创新、集成式创新成为此轮新科技革命的一个重要趋势。

（二）人工智能引领新科技革命

《福布斯》杂志2022年11月发布报告认为，人工智能、元宇宙、区块链、数字孪生、基因编辑、量子技术、机器人等技术有望在2023年重构世界面貌。中国科学院曾组织全国的专家总结了影响未来的12项关键性的前沿技术，包括数据编织、网络安全、隐私增强计算、云原生平台、决策智能、超自动化、人工智能工程化等，其中以人工智能为代表的类脑技术凭借着强大的产业变现能力，已经率先从众多的前沿技术创新中脱颖而出，成功地开启了建设地球智慧中枢的进程。目前，人工智能已经渗入各个领域，既是各行各业产业升级的重要工具，也将成为组织各行各业运行的智慧大脑，新科技革命将因人工智能的引

领，而呈现出爆发式的增长和突破。

（三）科技革命与产业变革紧密相连

新一轮科技革命与此前的蒸汽机革命、电力革命和信息革命一样，同样表现出科技创新拉动产业变革的特征。本轮科技革命是基于互联网技术和人工智能技术的革命，数据作为生产要素加入了生产函数，深刻改变了科技创新的范式和生产制造的可能边界，科技创新和产业化变革同步创新、彼此拉动的特征更加明显。

数据驱动、场景创新是新技术应用的主要推手，现代产业体系建设也主要围绕着数字化、网络化、智能化和绿色化展开，全球产业链和价值链体系因此重塑。对此，习近平总书记早在2013年主持十八届中共中央政治局第九次集体学习时就强调："即将出现的新一轮科技革命和产业变革与我国加快转变经济发展方式形成历史性交汇，为我们实施创新驱动发展战略提供了难得的重大机遇。"2023年5月27日，习近平总书记在陕西调研时又进一步指出"要以科技创新为引领，加快传统产业高端化、智能化、绿色化升级改造，培育壮大战略性新兴产业，积极发展数字经济和现代服务业，加快构建具有智能化、绿色化、融合化特征和符合完整性、先进性、安全性要求的现代产业体系"。

（四）伦理问题凸显

值得引起高度重视的是人工智能引发的科技伦理问题。

首先，自2016年DeepMind推出的围棋机器人阿尔法狗（AlphaGo）战胜韩国棋院第三代掌门人李世石后，包括李世石在内的数名围棋大师辞职。围棋的产生本是为人类的休闲活动和精神娱乐服务的，将人工智能应用于此类人类的精神活动领域将直接消灭最体现人类智力的乐趣，影响人类“自由个性”的发展。马克思把“个人全面发展”视为共产主义社会的标志。从社会关系需要消灭阶级角度，将之作为消除分工岗位限制的前提；从自然关系趋向“自由个性”角度，将之作为劳动能力脑力化的结果。社会主义的科技伦理，应充分展现社会主义制度是实现人的全面发展的社会条件，前瞻性地对人工智能所能参与或涉及的领域进行划定，避免从长远来看对人的全面发展、自由个性充分释放产生负面影响的做法。

其次，脑科学的研究还属起步阶段，模拟人类大脑到真正实现强人工智能之间，还存在较大差距。即使是能够完全模拟出人脑的神经元、神经突触等结构，也未必能够得到“强人工智能”。目前，仅仅通过人类大脑结构的简单模拟，并不能真正破解人脑的复杂功能。

再次，法律、科学伦理也在阻止强人工智能的研究，就像此前科学界反对克隆人一样，人工智能的伦理问题也非常突出。一个超越了人类智力的机器，再具备独立的心智和情感，是否会“甘心”服务人类？这种担忧不解决，

以人工智能革命为核心的新一轮科技革命，必然会因伦理问题而停滞。

二、新科技革命的驱动力量

（一）经济危机驱动新技术的应用与创新

科技创新与产业变革之间并不是一一对应关系，以科技成果转化率来衡量，2020 年西方发达国家的科技成果转化率已高达 60%—70%，而中国的科技成果转化率只有 30%。[1] 可见大量的科技创新成果往往会被束之高阁，并不能真正投入产业应用。

同时，产业的变革也并不一定由最新的技术创新驱动，真正能够助力企业成功的技术，往往是那些相对过时，却更贴合市场需求的技术创新，而非最前沿的技术突破。典型的案例就是 20 世纪 80 年代的操作系统竞争，当年 IBM 的 OS 系列操作系统性能更为强大、更为先进，却意外在市场竞争中败给了采用兼容模式的微软 Windows 操作系统。之所以会出现这样的错配现象，主要原因是技术创新只有得到产业应用的加持时，才能形成“技术创新—产业升级”的良性循环。而如果企业能够轻松获得外部市场的超额收益时，企业

[1] 孙颖妮：《中国科技成果转化率仅 30%，到底难在哪里?》，《财经》2022 年第 6 期，第 1—5 页。

就没有动力去应用最新的技术，追求安稳、确保盈利的动机会占据上风，一如20世纪80年代的IBM公司。

因此，将科技革命转化为产业变革，往往需要经济危机来扫清障碍，淘汰大量的低效企业，为市场的发展腾出宝贵的空间和资源。经济危机实现市场出清的同时，也会逼迫危机边缘的企业放手一搏，去应用最新的科技成果，从而开启产业变革的新赛道。

2008年和2018年，美国先后发生了两场金融危机，虽然由于各国央行的及时干预和联合行动，并没有演变为全球性的经济危机，但还是引发了全球化的全面退潮，各方因此都面临着存量竞争的巨大压力。在这样的情况下，全球科技创新进入了空前密集活跃的时期，试图以此重构全球创新版图和全球经济结构。

（二）人类的好奇天性持续推动科技创新

科技的创新、产业的变革都源自人类改造自然的愿望。正是因为人类对美好生活的不断追求，才有了源源不断的科技创新和产业变革。具体而言，人类的需求可以用马斯洛的需求层次理论来概括，从最底层的生存需要、生理满足，到最高层次的自我实现，共分为五个层次。需求的满足由低到高，逐步实现。对大多数人而言，通常只有低层次的需求满足之后，才会随之产生更高层次的需求。

此前的数次科技革命和产业变革，都可以视为人类在

努力解决物资短缺的困境。直到20世纪下半叶，人类才通过规模化、标准化的工业化道路，终于在信息化时代迎来了消费主义的大爆发。面对物质财富的丰富，人类开始追求更高层次的精神幸福，同时也对遥远的太空和更微观的量子世界产生兴趣，人类好奇的天性需要更大的满足。人类对更高层次需求的不懈追求，推动了人类持续不断的创新，科技革命是持续创新的时间累积和必然结果。

（三）互联网革命为新科技革命打下了坚实基础

具体到本次新科技革命，其突然兴起还得益于上一轮的互联网革命。互联网革命不仅为本轮科技革命提供了网络支撑，而且为新科技革命打下了产业基础，诸如半导体产业、人工智能产业都建立在此前信息技术革命构建的产业基础之上。

更为重要的是互联网的普及，使得信息获取更加便捷，知识交流的成本大幅下降，人才得以喷涌而出，这是本轮新科技革命的人才基础。而计算机技术和数据分析技术的成熟，使得人类有了更好的创新工具，创新不再过度依靠天才的灵光乍现，转而成为一个系统性的工程，更加依赖完整的创新体系和制度激励。正是互联网革命催化出来的这种组织化、体系化的研发模式，使得新科技革命的可预见性大幅提升。

（四）信息技术自身呈现指数级的增长

与此同时，信息技术的本身并没有像以前往的技术创

新一样，从兴盛走向衰落，走完技术创新到价格竞争的产品生命周期。信息技术在21世纪借助大数据技术、芯片技术和高速通信技术的成熟，演化升级为强人工智能技术，焕发了新的生命力，并改写了人类的发展方向。

与此同时，信息技术内部的各个关键技术也在快速发展。一是算力呈现指数级增长。芯片技术的进步使得海量数据的处理不再困难，而超级计算机的快速发展，进一步提升了算力水平，为大语言模型、人类基因测序这些需要大规模运算的行业创造了发展条件。二是泛在操作系统的成熟。计算机操作系统经过60余年的发展，在21世纪迎来了全新的泛在操作系统，其可以为不同的设备提供统一的界面管理，人类因此有条件开启万物互联的世界。三是软件重新定义世界。信息技术进化的另一个重要表现是软件的大发展，软件开始发挥主导作用，硬件由软件来定义其功能。软件将接管能源网络、无人工厂，人类会逐步进入一个“黑箱”时代，人类发出各种输入指令后，并不一定清楚后台软件是如何运行的，而只关心输出结果是什么。

信息技术的指数级增长，为新科技革命积累了技术成果和产业基础，人工智能、航天科技、生命科学等新科技革命的突破方向，都已经离不开信息技术的支撑，信息技术深度融入了新科技革命的各个领域，并成为其中重要的组成部分。

（五）人工智能引领新一轮科技革命

习近平总书记高度重视科技创新，多次为人工智能发展把脉定向。2018 年 10 月 31 日，习近平总书记在十九届中央政治局第九次集体学习时讲道："人工智能是引领这一轮科技革命和产业变革的战略性技术，具有溢出带动性很强的'头雁'效应。在移动互联网、大数据、超级计算、传感网、脑科学等新理论新技术的驱动下，人工智能加速发展，呈现出深度学习、跨界融合、人机协同、群智开放、自主操控等新特征，正在对经济发展、社会进步、国际政治经济格局等方面产生重大而深远的影响。"又在 2018 年给世界人工智能大会的贺信中深刻指出："新一代人工智能正在全球范围内蓬勃兴起，为经济社会发展注入了新动能，正在深刻改变人们的生产生活方式。"

（六）国家和政府成为新科技革命的重要推手

与以往的科技革命和产业变革不同，在新一轮的科技革命中，各国政府发挥了重要作用，不但是吹鼓手、引导人，甚至直接下场充当运动员。连美国这样一直标榜小政府的国家也撕下面具，直接开启了芯片战争。同时，还大手笔设立"电子复兴计划"（ERI）、国家人工智能研究与发展战略计划等国家级的发展规划，政府直接参与和资助相关项目，"不干预"市场的谎言就此戳破。

与此同时，欧洲国家、日本、中国等也高度重视国家

力量对科技创新的引领作用，这是以往科技革命和产业变革中十分罕见的现象。可能的原因，一是科技革命需要巨额的资金投入。仅凭企业的力量难以支撑。政府、企业、科研机构合作的创新，强调扩展公私合作伙伴关系，是新一轮科技革命的鲜明特色。二是新科技革命成为大国博弈的主战场。科技革命的巨大潜力为各国所认同，新科技革命带动的产业变革将决定未来世界的发展格局。因此，各国均高度重视，并积极参与。在某种意义上，新科技革命开启了“新时代的军备竞赛”。三是中国客观上需要新科技革命的发展机遇。改革开放以来，中国取得了令世界瞩目的发展成就，成功融入了全球生产和贸易体系是重要经验。中美贸易战以来，伴随着“去全球化”浪潮，中国面临产能过剩的危机，廉价劳动力的优势不复存在，客观上需要建立新的发展模式。拥抱新科技革命顺理成章，正当其时。

三、新科技革命的主攻方向

随着高性能计算、大数据和深度学习技术的快速发展，人工智能迎来第三次发展浪潮。在全球各国的积极推动下，人工智能持续向各个领域渗透与融合，涌现出一系列的新技术、新业态、新模式。市场对人工智能的投资回报也空前高涨，人工智能凭借技术前景和应用潜力，超越新科技

革命的其他领域，率先成为新科技革命的主攻方向。

人工智能的这种优势，可以概括为使能技术优势和应用场景优势，本书从科技革命的演进历程出发，重点聚焦人工智能的使能技术与应用场景两个层面，基于论文、专利、人才和商业数据，以及对专家的访谈等多方数据源，多维度分析人工智能作为新科技革命主攻方向的必然性、可行性以及现实情况，并以此为基础分析上海科创中心如何抓住其中的发展机遇，推动上海早日建成具有国际影响力的全球科技创新中心。

为此，本节我们将从演进历程、技术趋势、产业竞争三个视角，分别探讨人工智能作为新科技革命主攻方向的必然性与现实性。其中，首先对新科技革命的演进历程进行简单回顾，从科技演进周期的视角讨论人工智能技术崛起的必然性。其次，从算力建设、强人工智能、使能技术三个角度进一步论证人工智能作为技术主攻方向的可行性。再次，从产业趋势的角度，对市场现状、投资趋势、产业渗透率和应用场景进行分析、总结，并对人工智能的未来影响进行展望。

（一）从演进历程看科技革命的主攻方向

1. 人工智能是新科技革命的主攻方向

如前所述，历次科技革命和产业变革都可以概括为人类肢体功能的延伸，而新一轮科技革命的鲜明特色是人脑功能的延伸，人工智能凭借日益成熟的应用技术和迅速成

型的产业规模，事实上成为新一轮科技革命和产业变革的先导产业和主要现象。对此，许多国家相继出台了各自的人工智能发展战略，全球科技巨头们也纷纷把人工智能作为后移动时代的战略支点，努力在云端建立人工智能服务的生态系统。与此同时，传统制造业也在加速应用人工智能技术，推动新旧动能的转换，人工智能时代已加速来临。

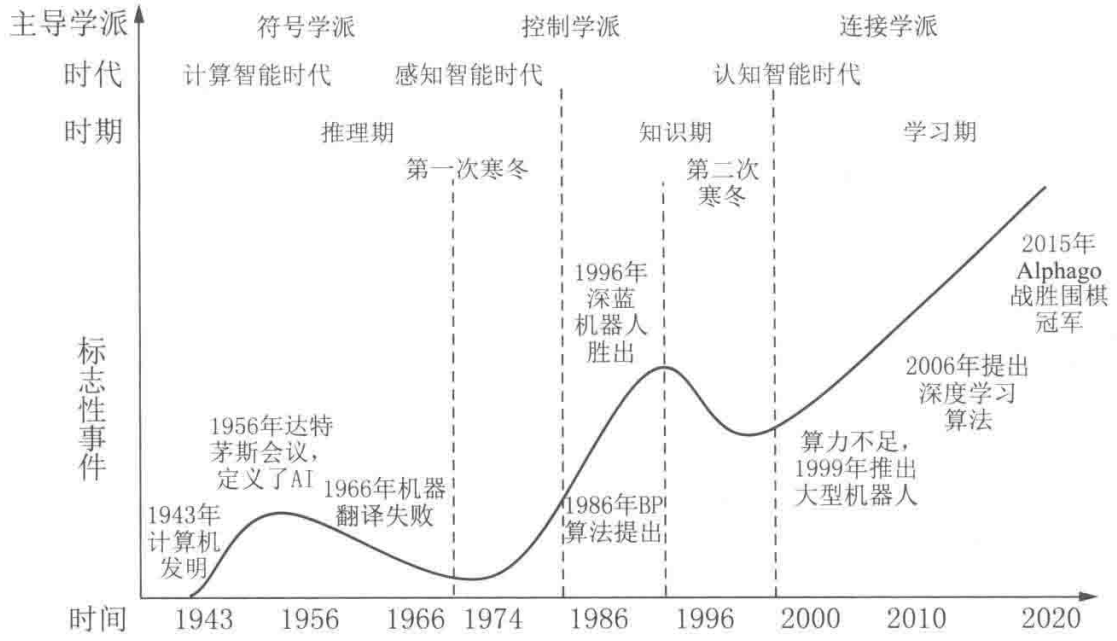
所谓智能，是指人类感知、思维记忆、学习与自适应、决策等方面的能力。人工智能（AI，Artificial Intelligence）是研究、开发用于模拟、延伸和扩展人类智能的理论、方法、技术及应用系统的一门科学技术，现在通常泛指为通过计算机、生命科学等技术手段，使机器具备感知、分析、推理和决策等人类智能的产品和相关服务。

机器代替人类认知、分析、识别和决策等功能，本质是机器对人类思维过程和人类意识的近似模拟。从 1956 年“人工智能”概念正式提出算起，在半个多世纪的发展历程中，人工智能已经经历了三次高峰和两次波谷，近年来，得益于计算能力、算法和数据三大核心要件的突破性发展，叠加已有的互联网优势，人工智能又一次迎来革命性的进步，成为新科技革命的主攻方向。

2. 人工智能的推理时代

人工智能的概念公认是起源于 1956 年的达特茅斯会议，这次会议历时一个月，将人工智能的思想带入学术殿

堂，并奠定了人工智能发展的学术框架。从 1956 年到 2023 年，近 70 年来人工智能的发展并非一帆风顺，经历了三个快速发展的时代和两个低潮期，前者包括推理时代、知识时代，以及机器学习时代，最终在 2010 年前后深度学习算法取得突破之后，应用才开始明显向好，并在 2017 年开始进入爆发期，进入了目前机器学习时代。



人工智能发展的五个时期

1956 达特茅斯会议开启了人工智能的第一次大发展，人工智能迎来了首个黄金期。这一时期的发展成果主要是机器具备了逻辑推理能力。以图灵测试为标志，产生了包括几何问题证明器、解题器等产品，而且美国军方在这一阶段给予了大量的经费支持。但由于计算能力、算法模型以及数据量都难以满足人工智能模型的算力需求，人工智能产品缺乏基本的常识和推理能力，只能完成待解决问题

的最简单环节。识别人脸等人类最基本的技能，对于机器来说都是巨大的挑战。这导致当时的人工智能无论是军用工程，还是商业项目都显得非常鸡肋，资本于是纷纷抛弃了人工智能的研究，相关研发经费被大幅压缩，军方也终止了对研究机构的支持，人工智能项目陷入停滞，进入1974—1980年的第一次低谷。

3. 人工智能的知识时代

这个阶段的重点是机器的知识学习，试图打造专家系统，将人类的知识传授给（训练）机器。当时很多大公司如DEC、杜邦等都开始部署专家系统，以降低决策成本，避免决策失误。在此期间，日本启动了第五代计算机的发展计划，投资570亿日元开发“采用平行架构的拥有人工智能的革命性的电脑”。英美两国军方也同样支持了类似的专家系统研发。中国科学院也计划过类似的研究项目，但由于经费短缺，未能付诸实施。

知识时代的研究成果以专家系统为代表。随着研究深入，基于知识库和推理的专家系统始终无法解决升级困难和稳定性不足的问题。专家系统需要海量的数据进行训练，虽然当时大规模集成电路已经出现，算力有所改善，但还是难以形成强有力的支持，而且维护成本居高不下，造成了商业化场景的匮乏，最终不得不下马。以日本宣布五代机研发失败为标志，英国、美国等也因研究成果不如人意，开始减少研发资金投入。人工智能的知识时代除了培养出

较多的行业人才外，并无多少实质性的建树。

4. 机器学习时代

这一时代从 2000 年开始直至现在。这个阶段的重点是通过大数据分析，以自动学习实现智能化。近年来，算力、算法以及数据都迎来了显著提升，奠定了行业应用的基础。深度学习算法、强大的计算能力和大数据分析等关键技术的创新式突破，促进了人工智能技术的空前发展，尤其是 2010 年后自然语言识别技术的实用化，带动整个人工智能行业进入实用化阶段。这个时代被业界称为“新一代人工智能”或者“人工智能 2.0 时代”。

各个时代产生了众多的人工智能思想，通常被冠名为不同的人工智能学派，具体见下表：

人工智能的学派

人工智能学派	主要思想	代表性方法
联结主义	利用数学模型来研究人类认知，用神经元连接机制实现人工智能	神经网络 SVM 等
符号主义	认知就是通过对有意义的表示符号进行推导计算，并将学习视为逆向演绎，主张用显性公理和逻辑体系搭建人工智能系统	专家系统 知识图谱 决策树等
演化主义	对生物进化进行模拟，使用遗传算法和遗传编程	遗传算法等
贝叶斯主义	使用概率规则及其依赖关系进行推理	相互 贝叶斯等
行为主义	以控制论及感知—控制系统原理模拟行为以复现人类智能	强化学习等

资料来源：网络

（二）从技术趋势看人工智能的主攻领域

技术变革是创新发展的底层驱动力，技术的演进趋势及其产业化进程，是衡量科技革命主攻方向的重要依据。过去四十多年，信息技术革命极大地推动了社会进步和产业发展，是上一轮科技革命的核心驱动和主导力量。根据“技术进步—产业创新—社会发展”的递进路线，我们可以根据人工智能从技术创新到产业应用的发展路径，来判定新科技革命的趋势，并据此制定人工智能的发展战略。

21 世纪前二十年，在超级计算机、GPU 芯片、并行计算、大数据、深度学习算法和类脑芯片等技术的推动下，人类社会先后迎来了互联网时代、大数据时代和人工智能时代。根据信息社会演进的历史经验，我们认为在人工智能发展的不同阶段，驱动力也各有侧重。

人工智能发展的驱动力

根据驱动力的变化，我们将人工智能的发展划分为三个阶段：技术驱动阶段、数据驱动阶段和场景驱动阶段。其中技术驱动阶段集中诞生了基础理论、基本规则和基本开发工具。在此阶段，算法和计算力对人工智能的发展起到主要推动作用，其中计算力主要包含芯片、超级计算机、云计算三个维度。当前，我们正处于人工智能变革的技术驱动阶段，技术能力和技术趋势是判断科技创新主攻方向的重要依据。

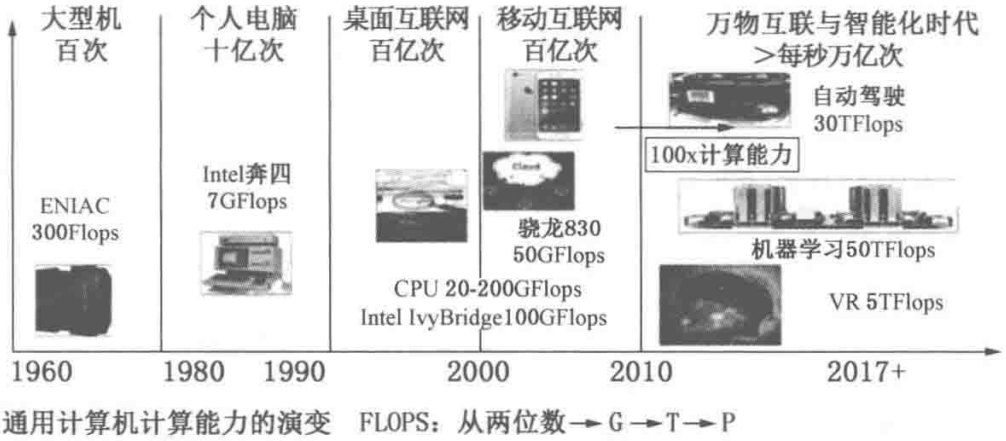
1. 算力明显提升，计算成本显著下降

人工智能硬件、算法及应用层均取得突破，行业进入爆发期：人工智能在60多年的发展过程中，经历了两次高潮和低谷，2017年以后再次进入爆发期。此次爆发主要源于人工智能硬件、算法及应用层三大要件的全面提升：一是算力提升明显，计算成本显著下降。二是算法取得重大突破，深度学习算法得到广泛应用。三是海量数据为算法实现和优化提供支持。技术走向全面，机器学习平台、多模态计算、多模型数据、计算能力的提升明显。除了CPU之外，相继出现了GPU、TPU、NPU、FPGA等人工智能专用芯片，而且算力基础设施的完善，也显著提高了计算能力，长期困扰人工智能发展的算力不足问题得到明显缓解。

（1）GPU芯片显著提升了人工智能的计算能力

CPU芯片以控制电路见长，是“全能型”选手，可以

完成综合性强、复杂程度高的任务，但是每次能够同时完成的任务量有限，因此在人工智能这种需要大规模并行，且重复性任务多的场景中，效果不尽如人意。此时 GPU 横空出现，表现明显好于 CPU。GPU 以前主要是用于图形计算，主要特点是重复计算的效率奇高。由于在计算图形点位变化的时候，需要进行大量的重复性向量计算，因此，对于人工智能这种需要大量并行计算、重复性运算的场景，GPU 的效率明显高于 CPU。以英伟达的 GPU TITAN X 为例，该产品在深度学习中所需训练时间只有 CPU 的 1/10 不到。



人工智能相关的通用计算能力演变

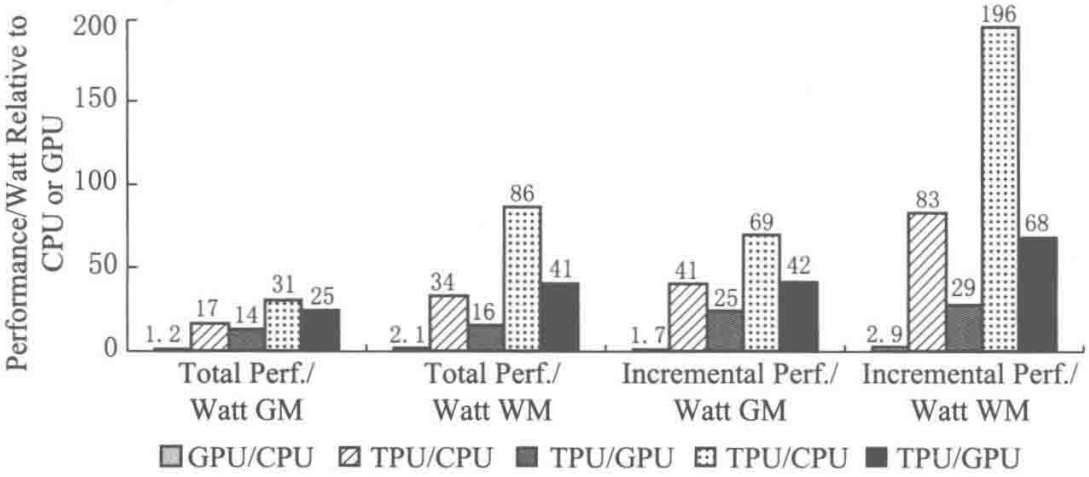
资料来源：中国信通院[1]

（2）TPU 芯片（张力处理芯片）的性能功耗比进一步提升

TPU 是谷歌专门设计的芯片单元，以进一步提升人工

[1] 备注：Flops 是衡量计算机计算能力的量，是 Floating-point Operations Per Second 每秒所执行的浮点运算次数的英文缩写。Gflops 表示每秒 10 亿次（109）浮点运算，Tflops 表示每秒 1 万亿次（1012）浮点运算，Pflops 表示每秒 1000 万亿次（1015）浮点运算，Eflops 表示每秒百亿亿次（1018）浮点运算。

智能芯片计算能力，同时大幅降低功耗。该单元2016年5月正式发布，此前已提前用于阿尔法狗与李世石的围棋对战。TPU被称为人工智能专用芯片，其针对TensorFlow等机器学习平台专门打造，可以在相同时间内处理更复杂、更强大的机器学习模型。谷歌通过数据中心测试显示，TPU平均比主流的GPU或CPU快15—30倍，性能功耗比（TFOPS/Watt）高出约30—80倍。



TPU、GPU、CPU和改进的TPU的性能对比

资料来源：In-Datacenter Performance Analysis of a Tensor Processing Unit、平安证券研究所[1]

（3）人工智能的性价比持续提升

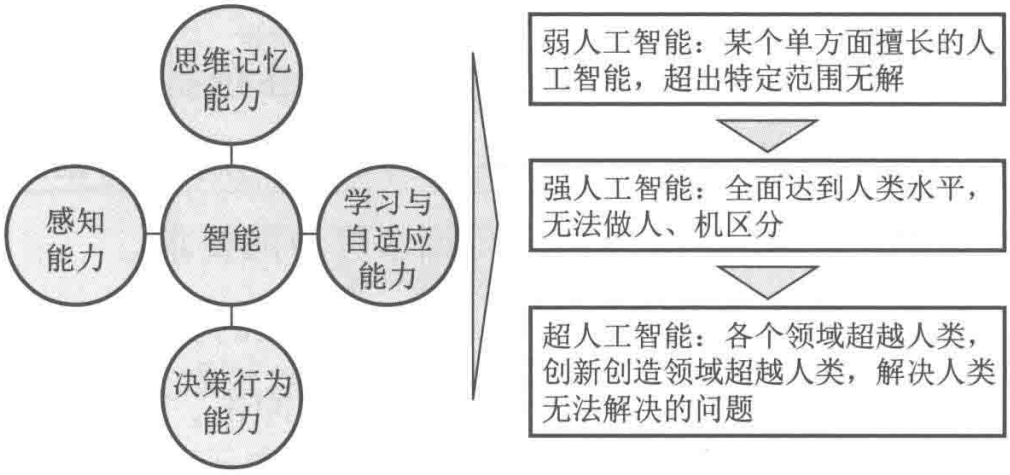
一直以来，人们通常以1000美元可购买的计算能力来衡量人工智能的性价比，其中CPS（每秒计算次数）是衡

[1] 备注：GPU（最左侧浅色）和TPU（斜杠）相对于CPU的能效比（功耗/瓦，TDP），以及TPU相对于GPU的能效比。TPU'是改进的TPU。点状代表改进的TPU与CPU对比的性能，深色代表改进的TPU与GPU对比性能。总和数据（Total）包含了主机服务器的功耗，增量数据（Incremental）则不包含主机服务器的功耗。GM和WM是几何与加权平均数据。

量计算能力的标准单元。根据人工智能专家库兹维尔的研究，1985 年 1000 美元只能买到人脑算力的万亿分之一，到 2015 年 1000 美元已可购买超过老鼠大脑的算力，该算力水平相当于人脑（1016 CPS）的千分之一，性价比是 1995 年的百万倍。根据库兹维尔的模型，业界预计到 2025 年，1000 美元可以买到相当于人脑的计算能力，算力的短板将不复存在。

2. 强人工智能标志着科技革命进入新阶段

学术界按照人工智能对人类能力的模拟程度，分为弱人工智能、强人工智能和超人工智能，三者之间在程度上是递进关系。



智能的构成以及人工智能分级

“弱”人工智能强调的是人工智能的工具性。其主要思路是借鉴人类的某些智能行为，减轻人类在某些领域智力劳动的负担。“强”人工智能甚至是“超”人工智能强调的

是“人造智能”，研发出具有心智和意识，并能够按照其心智和意识进行行动的人造物或者机器。

“强”人工智能和“弱”人工智能的争论从人工智能概念诞生之日起就已存在。争议的焦点主要有两个方面：一是什么是强（弱）人工智能，区分的标准是什么？如何实现？二是要不要发展强人工智能？这两个方面的争论一直影响着人工智能的发展，我们将在第三章的（三）场景产业与生成式人工智能中，详细讨论以 GPT 为代表的强人工智能。

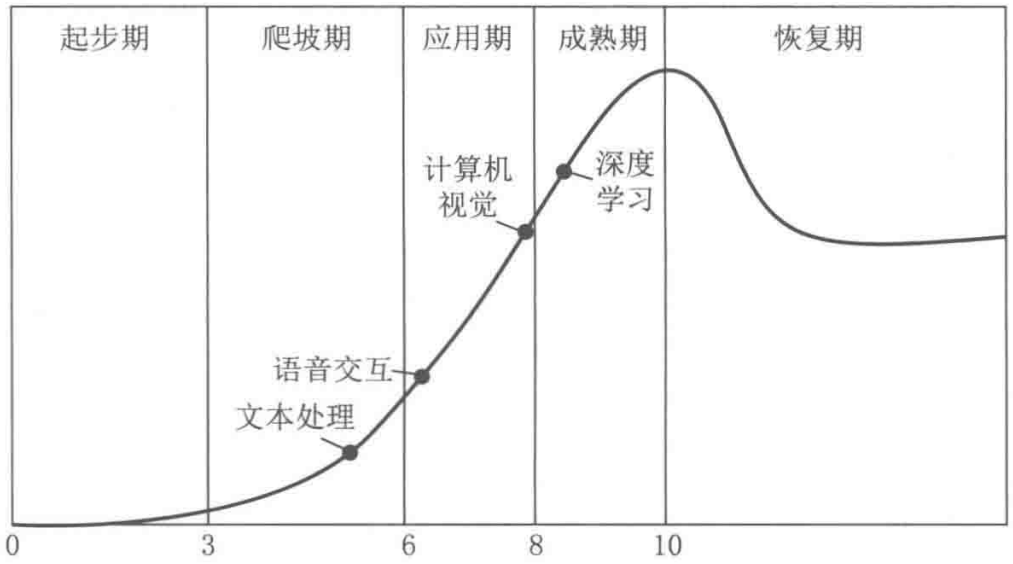
机器学习、深度学习作为强人工智能最基础最关键的核心技术，驱动图像、视频、语音、语义、知识图谱等技术的应用，未来还将向融合语音、语义、视频图像等多模态计算发展。而从技术到应用，并真正在企业生产和社会生活中发挥效能，则需要行业参与者共同解决数据资源、硬件适配、云端协同、工程化等诸多环节的要求。

目前，具有通用性、基础性和使能性的强人工智能技术已经出现，正对各行各业带来极大的冲击，创造出新的产业部门、商业业态和价值链体系。新旧主导产业、领军企业的更替，已经开始改写国家和企业的竞争格局，这是科技革命到产业变革的基本特征。

3. 人工智能技术达到了实用化标准

相比于其他技术方向的突破，人工智能革命最重要的

还是要将技术与应用场景相结合，无法落地的技术很难得到资本青睐和市场认可，而有闭环、垄断性的数据，并且其技术能够与实际应用场景结合的公司将有望产生难以被替代的商业价值。随着 AI 支撑技术的不断发展，人工智能将持续拓展更多的应用场景，应用场景将不断丰富，并会反过来驱动支撑技术的持续发展，人工智能的整体市场规模将继续扩大。



人工智能使能技术发展曲线

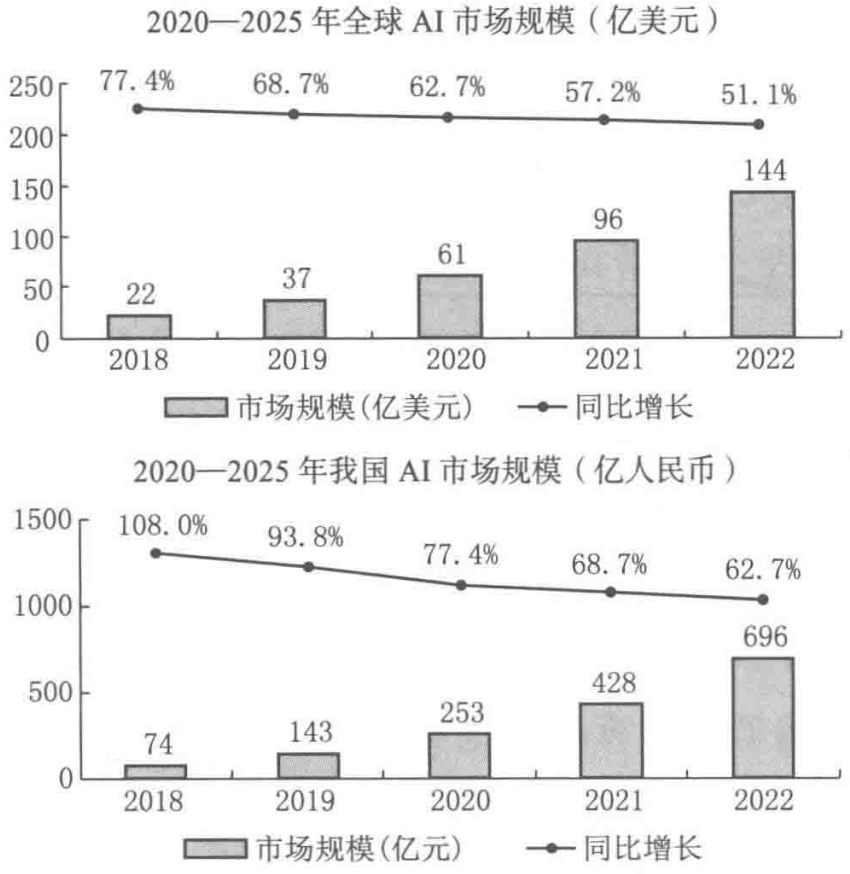
以使能技术就绪度指数为标准，目前深度学习就绪度达到 8.3，已处于成熟期，并越来越多地应用到各种实际场景中。计算机视觉就绪度次之，语音交互和文本处理的就绪度相对偏低。计算机视觉和语音交互尚处在技术应用初期，并开始尝试应用于不同的场景，而文本处理仍处于技术爬坡期，期待 ChatGPT 能够实现创新突破。

（三）从产业趋势看人工智能革命

人工智能作为新一轮产业变革的核心驱动力，将进一步释放历次科技革命和产业变革积蓄的巨大能量，并创造新的强大引擎。受益于互联网的发展，中国的人工智能技术已经走在世界前列，人工智能赋能各行各业的进程正在加速，涌现了一大批智能制造、智能教育、智慧农业的经典案例。

1. 人工智能市场规模持续扩大

根据 Statista 的数据，2025 年全球人工智能市场的规模预计达到 368 亿美元。保守估计 2022 年中国的人工智能市场应接近 700 亿元的规模。

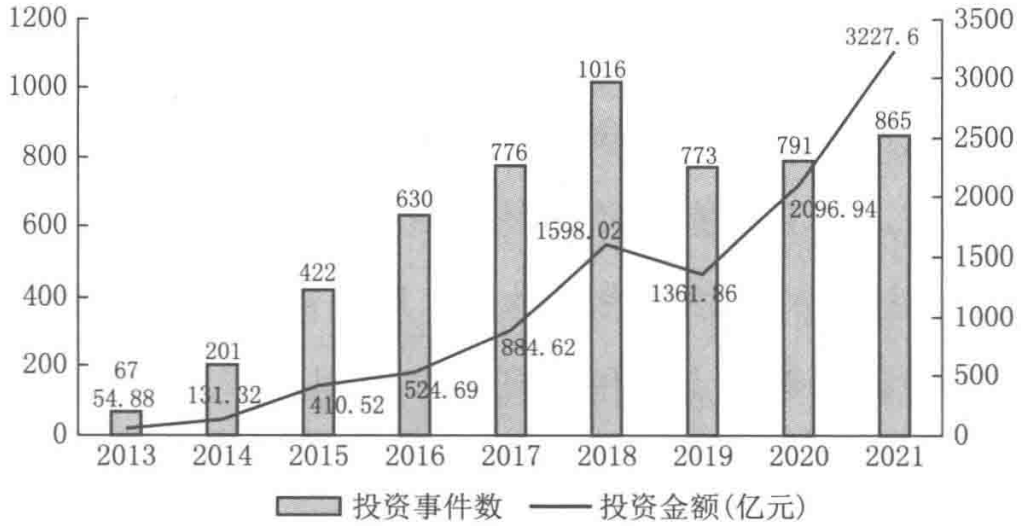


人工智能产业的市场规模

资料来源：Statista，36 氪研究院

2. 各路资本热捧人工智能市场

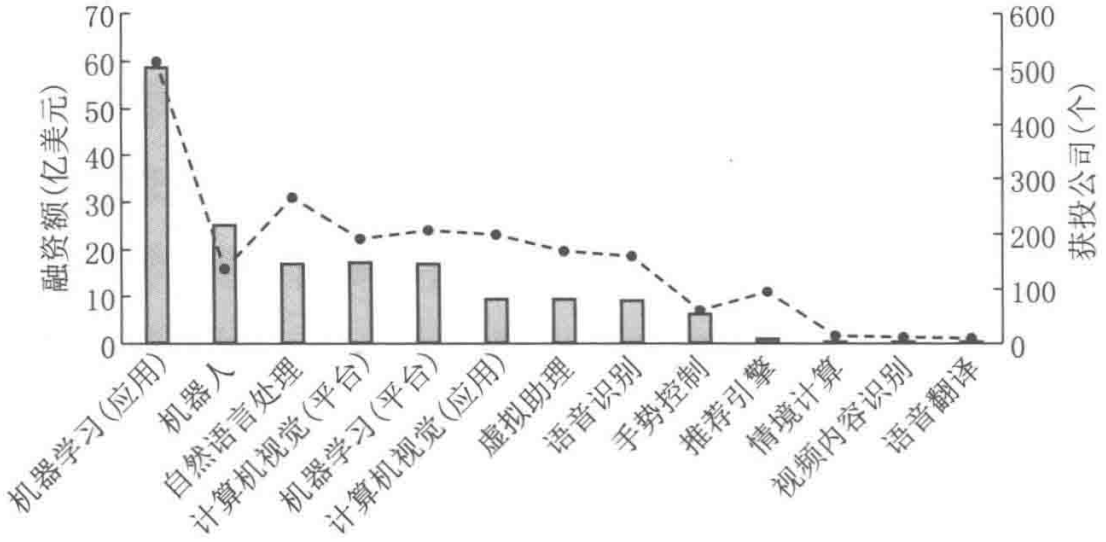
人工智能的发展前景受到了资本市场的高度青睐，近些年来，尽管全球经济危机的乌云密布，但 AI 领域投融资热度不减，全球人工智能企业的投融资规模持续扩大。根据前瞻产业研究院的数据，2013—2021 年人工智能行业的融资额年复合增长率为 66.4%，未来有望进一步迎来爆发式的增长。



2013—2021 年全球人工智能投融资情况

资料来源：IT 桔子、前瞻产业研究院

根据国际数据分析机构 IDC 预计，在 2022—2026 年期间，全球 AI 市场将继续以 18.6% 的复合年增长率增长，2026 年有望达到 9000 亿美元的规模。且人工智能企业的创业热潮逐渐趋向理性，更注重核心技术和应用场景的落地。



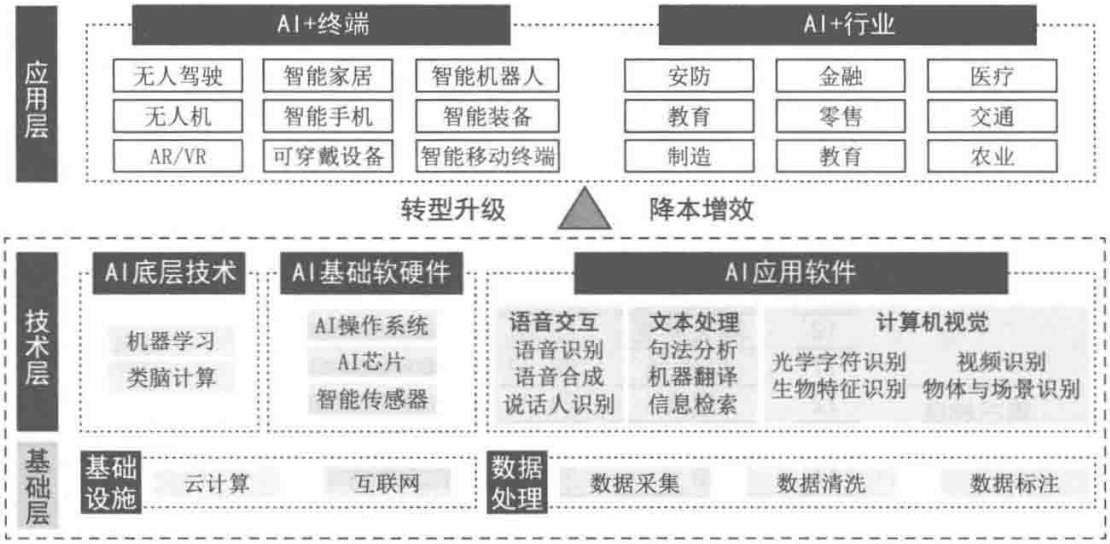
全球人工智能各细分领域获投金额 & 获投公司数目

资料来源：Venture Scanner

3. 人工智能的产业渗透率显著提升

当前，人工智能已经发展为新一代的通用技术，正加快向经济社会的各领域渗透融合。从语音控制、人脸识别、光学字符识别等单点应用，到生产流程、产品、信息消费和服务业的智能化、高附加值转型，人工智能在医疗、金融、安防、教育、交通、物流等多个领域开枝散叶，实现了新业态、新模式和新产品的突破式应用，成为推动经济社会发展的强大引擎。

展望未来，人工智能的产业化将进一步提速，在需求拉动和市场驱动的双重效应下，人工智能技术的渗透率将进一步提高，将从生产制造领域进一步融合社会服务领域，市场生态将高度整合。



人工智能与产业升级

4. 人工智能产业率先进入了爆发期

当前，欧盟、日本等已出台完整的数字经济政策和人工智能发展战略。中国人工智能的“AI+”战略正在稳步推进，人工智能产业正由萌芽期进入快速爆发的成长期。

（1）人工智能的产业体系初步成型

人工智能产业的发展离不开国家战略的扶持，但更得益于核心企业的引领发展，中国和美国概不例外。美国的头部大数据企业和人工智能企业主要有谷歌（Google）、苹果（Apple）、亚马逊（Amazon）、特斯拉（Tesla），已经建立了从底层算法到人工智能基础设施，再到终端场景应用的完善体系。

在中国，大企业同样充当了人工智能产业的领头羊，发挥着前沿创新的巨大作用。中美两国的共同优势是拥有人工智能产业的全部要素，大数据、搜索引擎、算力基础

	排名	国家	得分
第一梯队	1	美国	59.43
	2	中国	50.14
第二梯队	3	韩国	37.29
	4	英国	34.28
	5	新加坡	31.11
	6	加拿大	28.92
	7	德国	27.89
	8	日本	27.05
	9	法国	26.91
	10	澳大利亚	26.81
	11	以色列	26.37
第三梯队	12	瑞典	24.36
	13	卢森堡	23.93
	14	荷兰	21.85
	15	比利时	21.61
	16	丹麦	21.06
	17	芬兰	20.80
	18	奥地利	19.86
	19	西班牙	17.31
	20	葡萄牙	17.16
	21	意大利	17.00
	22	爱尔兰	16.68
	23	捷克	15.50
	24	爱沙尼亚	15.14
第四梯队	25	印度	14.95
	26	俄罗斯	14.44
	27	斯洛文尼亚	14.34
	28	土耳其	13.85
	29	匈牙利	13.68
	30	波兰	13.63
	31	塞浦路斯	13.44
	32	马耳他	13.34
	33	希腊	13.12
	34	克罗地亚	13.03
	35	拉脱维亚	12.05
	36	斯洛伐克	12.02
	37	立陶宛	11.63
	38	沙特阿拉伯	11.31
	39	保加利亚	11.26
	40	巴西	11.03
	41	罗马尼亚	10.32
	42	南非	10.16
	43	越南	9.91
	44	阿根廷	8.31
	45	墨西哥	6.44
	46	印度尼西亚	5.18

部分国家人工智能创新指数排名

资料来源：中国科学技术信息研究所

设施、人工智能人才的产业发展要素基本齐全，未来人工智能产业的竞争也将主要在中美两国展开。相比而言，欧洲过度关注数据立法，试图率先占领数字经济和人工智能的道德和法律高地，其人工智能产业和技术的整体竞争力相对较弱。

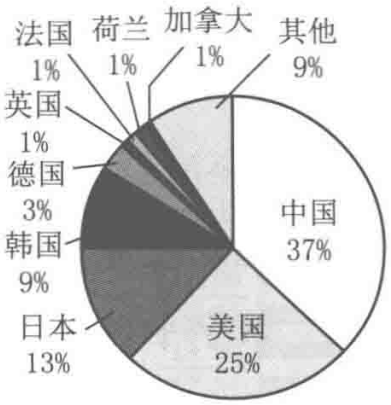
（2）多元创新的格局基本成型

除了引领发展的核心企业之外，人工智能领域还有大量的实验室、研究机构、数字平台和创业公司，构成了一个生机勃勃的人工智能产业生态圈，大企业也因此时刻充满危机感。因为，初创公司的数量成为体现各国人工智能竞争力和活跃度的一个重要指标。根据新华社数据，截止到2023年世界人工智能大会召开时，中国人工智能产业的核心产业规模达到5000亿元，企业数量超过4300家，中国人工智能产业的创新活力十足。

在国家竞争格局方面，根据中国科学技术信息研究所发布的《2021全球人工智能创新指数报告》显示，截止到2021年9月，美国有人工智能企业4664家。全球46个发展人工智能的国家可以分为四个梯队，其中美国的人工智能产业排名位居全球第一，遥遥领先世界各国，与中国一同位列全球第一梯队。韩国、英国、新加坡、加拿大、德国、日本、法国等9个国家各具竞争优势，呈现彼此竞争、你追我赶的态势，位列第二梯队。瑞典、卢森堡等13个国家在某

一细分领域内具有一定的优势，处于第三梯队。而印度、俄罗斯等 22 个国家则属于人工智能发展的第四梯队。

（3）人工智能专利相对集中



全球人工智能专利地域分布
资料来源：信通院、平安证券研究所

5. 人工智能对经济的贡献快速增长

随着人工智能在硬件、算法以及应用领域的不断发展，对各行各业渗透率的不断提升，人工智能引领全球第四次工业革命浪潮的作用已经凸显，人工智能在全球经济社会中的影响力将逐步累积。

（1）人工智能的经济拉动效应显著

据麦肯锡的研究报告，到 2030 年人工智能产业可能为全球贡献 13 万亿美元的 GDP 增量（相较于 2018 年），较 2018 年基础上增长 16%，年均增速 1.2 个百分点。人工智能产业对全球经济的拉动效果足以比肩人类历史上前三次通用技术革命（蒸汽机、电气化、信息化），因此也有研究

称：人工智能将引领“第四次工业革命”的兴起。

（2）人工智能的经济贡献还加速提升

当前人工智能技术仍处于实用化的早期阶段，远远没有达到高度成熟的状态。比如人工智能还不能达到人类思想的深度，行业的深度应用还有众多的技术瓶颈，从技术成熟到全球化应用仍需要较长时间。当前人工智能的学习成本居高不下，算力建设的前期投入非常高昂，许多大数据中心普遍处于亏损状态，企业应用和推广人工智能技术的速度也较慢。但我们相信，人工智能的时代已经来临，随着企业竞争的加剧，人工智能渗透率将爆发的奇点已然不远，对经济的贡献也将加速提升。麦肯锡公司对此预计，到 2025 年人工智能产业将进入加速发展的阶段，到 2030 年左右，人工智能对经济的贡献度将是 2019—2023 年阶段的 3 倍左右。

四、人工智能革命的深远影响

新科技革命是一次改变人类命运的革命，从学习模式、工作方式到消费模式，人工智能将彻底改变人类的生活理念和存续模式，人类文明很可能将进入“再生时代”。随着全球科技竞争的加剧，主要发达国家纷纷强化重大前沿技术部署，新一轮科技竞赛和产业竞争已经全面展开。新科

技革命将决定每一个国家和个体的未来，人工智能就是当前竞赛的主战场。由此，怎么评价人工智能的影响都不为过，当前的人工智能革命将为我们每一个人带来翻天覆地的变化。

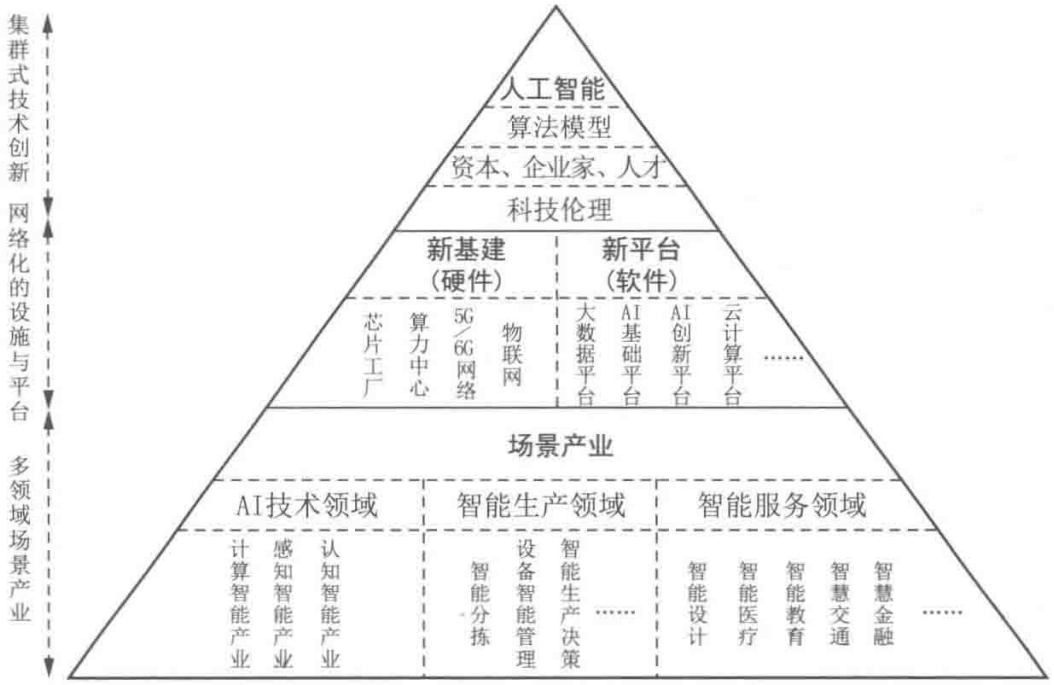
第三章

人工智能革命的金字塔框架图

一、金字塔框架图简介

（一）金字塔框架图的架构体系

新一轮科技革命的主攻方向是人工智能，可以概括为类脑技术革命。基于政府部门的视角，此轮人工智能的技术创新到产业变革，可以用金字塔框架图进行描述（详见下图）。



新科技革命——金字塔框架图

金字塔框架图的第一层塔尖部分，是作为新一轮科技革命主攻方向的人工智能技术，体现了新一轮科技革命“技术创新引领产业变革”的典型特征。以人工智能技术创新为引领的新科技革命，是“科技引领、需求拉动、供给创新、市场竞争”综合作用的产物，具有集群式创新、网络化协调的特点，是产业升级与组织创新彼此促进、协同发展的过程，需要科学家、企业家、工程师和金融家等各方力量共同推进。

金字塔框架图的第二个层次，是支撑类脑技术创新的各种软硬件体系，对标国家的政策文件，我们将之称为“新基建、新平台”。以人工智能为代表的类脑技术革命，其产业化、市场化的鲜明特征是网络经济、平台经济和智能经济，这些都建立在新型网络基础设施和数字设备的基础上，必须超前建设、超前部署。新型基础设施建设不仅具有明显的投资拉动效应，能够有力地促进经济增长，而且能够推动传统产业的智能化转型升级，开启新一轮的经济增长周期。因此，尽管目前人工智能技术的应用场景还不十分清晰，但新基建、新平台的建设依然受到各方的高度关注，各级地方政府更是积极参与其中，成为新基建的重要推手。综合分析全国各地已公布的数字经济发展规划，各地普遍将未来的智能经济理解为数字产业化、产业数字化和智慧城市，而对新基建、新平台的具体内容则普遍缺

少清晰的描述。这反映出我们对新基建、新平台缺少系统规划，更多局限于为具体应用场景搭建专用的数字平台，存在一定的盲目性。

金字塔框架图的第三层次，是基于新基建、新平台的各类场景应用。这一层次具体到各个细分行业的应用，是人工智能技术等高新技术在各个行业、各种环境下的场景应用，是当前各地积极引进、大力推动的具体项目。场景应用具有鲜明的需求驱动特征，而此轮科技革命的鲜明特征是供给创新，是技术创新引领产业变革，以及市场产品快速迭代更新反向拉动的持续技术创新。因此，以人工智能推动的供给侧技术创新和产业变革，如何与需求侧驱动的场景应用形成良性对接，是各级地方政府在推进新科技革命、发展人工智能产业时，需要重点考虑的问题。尤其需要关注的是应用场景开发具有碎片化的特征，零散的应用场景项目如果不能持续跟上智能技术、数字网络的更新速度，极易形成新的信息孤岛，最终变为低端产业、过时装备，这是各级政府尤其是基层政府和业务部门在推进人工智能应用时，需要注意的问题。

（二）金字塔框架图的结构关联

1. 驱动力转换的三个阶段

金字塔框架图是从政府视角来看待人工智能革命的，其框架体系依据人工智能发展的三阶段而构建，各层次分

别对应从技术创新到产业变革的动力转换。根据驱动力不同，可以把人工智能的发展划分为“技术驱动、数据驱动、场景驱动”三个阶段。三个阶段的创新内容彼此关联，互为因果。与之对应，人工智能的架构体系分为三个层次，自上而下分别为“人工智能、新基建新平台、场景产业”，详见金字塔框架图。

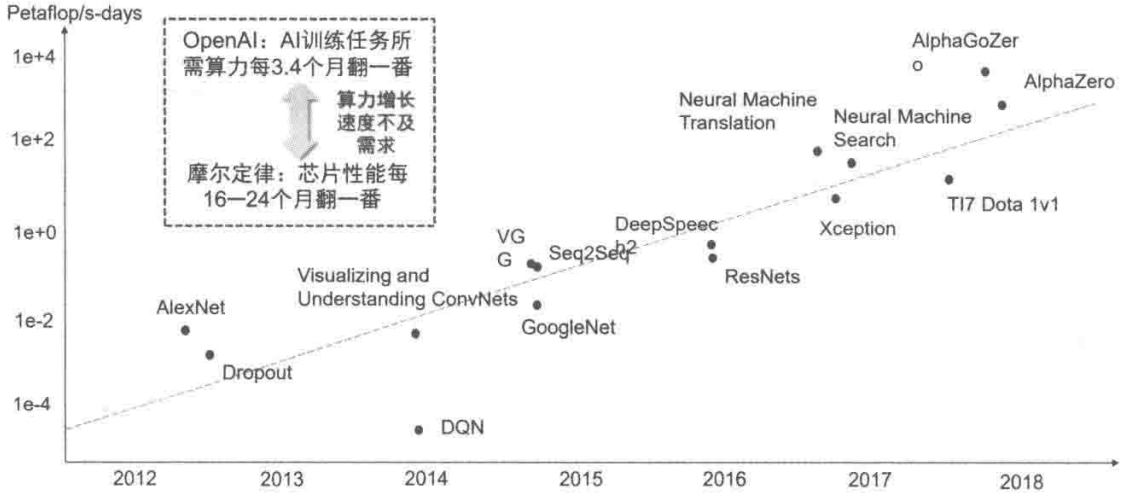
（1）技术驱动阶段

技术驱动阶段对应金字塔框架图中的塔尖部分，即人工智能部分。技术驱动阶段是人工智能的起始阶段，人工智能的基础理论、核心技术和基础工具在此阶段不断完善并走向成熟。在技术驱动阶段，算法的完善和算力的提升是工作的核心，对促进人工智能复兴厥功至伟。2010 年以来，长期困扰上一代智能专家系统的复杂模型求解问题，在新一代大数据模型和超级计算机的帮助下，算力瓶颈迎刃而解。以 2016 年 3 月阿尔法狗战胜围棋世界冠军李世石为标志，人工智能取得了又一个里程碑性质的胜利，强人工智能呼之欲出。技术的实质性突破，使整个人工智能产业重获新生，带动了相关的新基建、新平台的快速兴起。

（2）数据驱动阶段

数据驱动阶段对应金字塔框架图中的新基建、新平台部分。数据驱动阶段的核心内容是算法、算力和算据，可以简单概括为智能芯片、超级计算机、云计算。人工智能

进入数据驱动阶段后，其强大的性能迅速得到各行各业的青睐，用户群体迅速扩大，行业应用持续推出，海量数据喷涌而出并快速膨胀，新基建的需求由此引发，以满足海量数据传输、运算和存储的需求。这一阶段算力基础设施的短板始终存在，直至各种场景应用替代数据要素，成为企业利润的主要来源。



算力需求的增长趋势

资料来源：OpenAI 官网，中信证券研究部

（3）场景驱动阶段

场景驱动阶段对应金字塔框架图最底部的多领域场景应用技术（产业）。场景应用技术成为第三阶段的主要驱动力，源于技术创新固有的演进规律和扩散模式。通常核心技术取得突破后，会沿着“通用技术普及—专用技术研发”的路径，以雁阵形式不断向各个领域渗透与融合，推动“产品创新—工艺流程创新—市场模式创新—企业组织创新”循环的往复。虽然互联网的普及以及数字技术的创

新，极大地改变了企业的创新模式，尤其是网络协同创新、市场模式创新乃至人工智能的自我研发等，很大程度上改变了技术研发的生态环境和技术创新的应用模式，但“核心技术—通用技术—专用技术”的创新分类和推广顺序并没有完全改变。

具体而言，“场景”（context）一词由影视用语演化而来，始见于《即将到来的场景时代》一书，[1] 指移动互联网时代，用户、生产者、产品（或服务）三者之间，能够借助移动设备、社交媒体、大数据、传感器和定位系统等“场景五力”，人为构设并建立起集合和环境，进一步强化两两相关的紧密关联，通过网络协同为场景中的用户提供更好的体验，在此过程中实现价值变现。因此，场景驱动阶段不仅仅是人工智能技术向各行各业的渗透与融合，更是构建一个全新的虚拟场景空间，以全新的界面形式实现实用性功能和社会性功能。这种虚拟空间和情境体验的技术创新，也必然驱动各行各业发生翻天覆地的变化。

2. 从技术关联到产业关联

如上所述，从技术关联的视角来看，人工智能从原创技术的突破到支撑技术的扩张，再到各项应用技术向各行各业的持续渗透与逐步成熟，人工智能技术的原始创新如

[1] ［美］罗伯特·斯考伯、谢尔·斯雷尔：《即将到来的场景时代》，赵乾坤、周宝曜译，北京联合出版公司 2014 年版。

同水中涟漪一般，不断扩散、叠加升级。围绕核心的智能算法模型，构建起一个庞大技术关联体系，以此拉动轰轰烈烈的人工智能产业革命。

（1）算法产业化

技术驱动转化为产业拉动的过程，与技术的关联、扩散和成熟进程息息相关。与三层次的技术体系相呼应，人工智能的产业体系相应也可以划分为三个层次：核心技术产业（算法）、算力及算据产业、场景产业。其中，核心技术产业主要对应金字塔框架图中塔尖部分，体现了人工智能以算法模型为核心的先导产业特征。核心的先导产业通常包含算法技术本身形成的产业，以及围绕算法模型构建的人工智能核心技术产业，如与基础层、感知层和认知层直接相关的产业，具体我们将在本章第二节专题介绍。

（2）算力产业

人工智能产业体系的第二个层次可以概括为算力产业，与新基建、新平台相对应。算法、数据、算力是人工智能的三大要素，其中，算法和数据两要素都离不开算力的支撑，算力是人工智能产业化的必要前提。由于算力产业的门槛相对较低，又是推动数字经济发展的核心力量，因此受到了各方的高度关注，成为企业加速布局的重点。

所谓算力，就是计算能力。算力大小反映了数据信息的处理能力，也体现了人类在机器智能领域的发展水平。

2018 年以来，随着人工智能算法的飞速发展与不断成熟，大模型的参数量级已达到数千亿个。这种基于海量自然语言或多模态数据集训练的生成式 AI，其对算力的需求呈现指数级增长，导致算力成为人工智能发展的瓶颈。

根据 OpenAI 的报告数据，训练 ChatGPT-3175B 的模型，需要的算力高达 3640PF-days（即假如每秒计算一千万亿次，需要计算 3640 天），需要 7—8 个算力 500P 的数据中心才能支撑运行。以国内领先的贵阳大数据交易所为例，2023 年 4 月，其向社会公开了全国首个“算力资源专区”平台，目前接入算力为 29.7P，以此计算，需要 122 个贵阳大数据交易所平台才能支撑 ChatGPT3 的运行。与此同时，无人驾驶等智能设备的推出，也对人工智能提出了更高级别的实时反馈要求，进一步加剧了算力资源的短缺。

短缺意味着机遇，强化新基建、新平台建设，顺理成章成为热点。数据中心、5G 网络、芯片工厂，以及与之配套的高性能服务器、GPU 等相关新型算力基础设施，都得到资金热捧，行情看涨。

（3）场景产业

人工智能产业的第三个层次是场景产业，对应人工智能体系的应用技术层，是人工智能技术向各细分行业渗透与应用的产物。场景产业汇集了解决各行业具体问题的智能方案，既是人工智能的应用技术集成，也是不同应用场

景的产品集合。由于细分行业的庞大规模，以及人工智能的行业普适性，因此，人工智能赋能实体经济能够创造出庞大的市场，场景产业的规模也远超前两个层次，这使得人工智能的应用技术和场景产业得以成为一个独立的层次。

事实上，自人工智能技术由实验室走向大众那一刻，应用场景一词就成为热门词汇，不断被各级政府官员反复提及。应用场景由技术应用升格为场景产业，其背后体现的是技术驱动到产业驱动的转型。当大规模的新型基础设施投产运营后，场景应用就凭借更好的服务体验、更高的运营效率、更普遍的财富效应，取代新基建成为行业竞争的焦点，场景产业的发展开启了场景驱动的第三阶段。

数据驱动阶段与场景驱动阶段更多是一种学理区分，通常并没有一个清晰的判断标准和界限区分。事实上，在实务中两个阶段的发展内容，很多时候是相互交织在一起的，通常同步规划、同期推进。比如在建设新基建、新平台的同时，就需要对目标客户、应用场景、盈利模式等进行精准规划，而非盲目地先建再说。以大数据中心为例，国家级的大数据中心主要面向超大规模运算，会更多考虑数据、能耗、安全等因素，所以在数据驱动阶段就开始大力建设，是战略性的举措。而企业级的数据中心，会更多考虑市场需求、应用场景、定价标准、成本收益等因素，更多地基于场景应用和市场需求，主要在场景驱动阶段投产。

二、塔尖的人工智能

（一）逻辑架构

1. 人工智能技术的三个层次

在金字塔框架图中，塔尖部分是作为创新引领的人工智能技术，对整个人工智能革命起到了技术支撑和方向引领的作用。塔尖部分可以从技术逻辑上进一步分为三个层次：基础层、技术层和应用层。需要注意的是，整个金字塔框架图是从政府视角来看人工智能革命的，即新一轮科技革命先有了作为核心动力的人工智能技术，然后需要为人工智能体系建设新型基础设施，提供新的发展平台，再然后才能在新基建、新平台的基础上，利用人工智能技术发展出形形色色、丰富多彩的场景产业，推动整个社会进入数字经济、智能经济时代。即金字塔框架图从驱动力角度将人工智能革命分为“人工智能核心技术、新基建新平台、场景产业”三个层次。

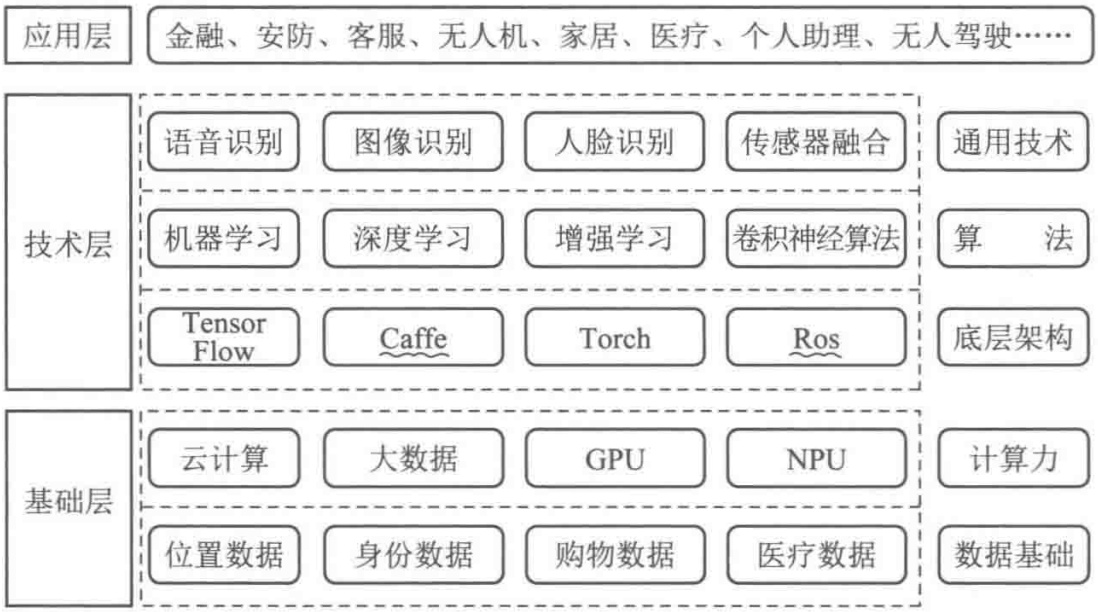
而如果进一步深入人工智能的塔尖内部，可以把创新引领的人工智能技术进一步分为“基础层、技术层、应用层”三个层次。两种分类讨论的是不同层次的概念，但由于人工智能具有高度的渗透性、融合性和协同性，人工智能技术与新基建、新平台、场景技术之间，彼此存在较多的技术交叉和产业重合。因此，虽然两种分类标准针对的

是不同层次的问题，但依然会存在许多概念交叉、产业重合的地方，这需要我们事先关注讨论的概念和范围。

之所以在塔尖部分回到技术分类的视角，是因为人工智能的发展涉及众多的前沿技术，各种新技术、新名词往往令人眼花缭乱、无所适从，各界对此至今尚未形成一个公认的标准体系。所以此处需要使用技术部门常用的框架来讨论。

2. 人工智能技术的逻辑架构

如下图所示，人工智能技术的逻辑架构分为三个层次：基础层、技术层和应用层。这种逻辑架构是按照“先硬件、后软件”的顺序分类，即先有基础层的数据和硬件技术，然后有更高一级技术层的架构、算法模型和通用技术，最后才发展出应用层的各种细分行业的场景应用技术。



人工智能技术的逻辑架构体系

资料来源：平安证券研究所（2019）

基础层的技术主要解决计算能力问题，具体有三个方面：一是海量数据如何获取、分类、存储，对应逻辑架构体系图中底部右侧的数据基础。相关的技术包括与数据相关的各类智能传感器技术、智能定位技术、大数据技术等。二是数据如何计算，对应图中底部右侧的计算力。相关的技术主要是各种云计算技术，以及能够计算海量大数据的超级计算机。而与之相对应的产业就是新基建、新平台中的各类数据中心。三是用什么设备来计算数据。与之相关的就是各类人工智能芯片，如 GPU、FPGA、NPU 等。

技术层主要是解决如何实现智能的问题。如果说基础层初步完成了大数据的归集、分类和简单运算，能够提供有限的弱人工智能，那么技术层的主要任务就是对数据要素进行深度加工，研发各类先进的算法模型，实现强人工智能。

为此，技术层应当包括如下内容：一是能够不断提升人工智能性能的各类算法模型。涉及诸如机器学习、深度学习、增强学习、卷积神经网络等，这些核心的算法技术形成产业化的项目，通常被称为算法产业化。二是实现强人工智能所需的各种感知技术。与基础层的智能传感器主要提供各类状态数据不同，这里的感知技术更多指语音识别、图像识别、人脸识别、传感融合（AR、VR）等高度拟人化的智能感知技术，使得智能机器和智能系统能够动态感知和识别外部复杂的世界。三是认知技术。认知部分可

以分为：自然语言处理、语义识别和知识图谱，是智能机器能够做出分析判断的基础。感知和认知技术是发展人工智能的关键性通用技术。

应用层也称为平台层，主要是基于平台模式展开的人工智能应用。需要注意的是，这里的应用主要针对人工智能的物理架构和技术体系而言，侧重人工智能的逻辑体系和技术应用。而金字塔框架图第二层次、第三层次讲的新基建、新平台和场景产业，则主要关注有形的物理实体和具体的场景应用。人工智能在基础层、技术层所研发的关键技术，如何才能转化为应用层的场景技术，关键是要处理好行业知识和场景接口的融合问题。对此，业界提出了元宇宙的解决方案，即建立一个虚拟的场景空间，通过各种场景接口实现与物理世界的一一对应。人类通过元宇宙的虚拟空间，借助人工智能的应用场景技术实现操作，从而大大减少直接面对复杂、高危、高难度、重复环境的工作风险，同时也大大提高实体资源的利用效率。因此，应用层的人工智能技术与场景产业是紧密相关的，前者是后者的软件内核，后者是前者的产业化落地。

（二）基础层

基础层的核心是“芯片、超级计算机和云计算”三驾马车，主要功能是为人工智能提供基础数据和算力支撑，对应新基建的算力产业化项目是芯片工厂、数据中心等。

其中，基础数据既包括消费数据、位置信息等实时数据，也包括行业数据、身份信息等历史数据。算力支撑则主要由各类人工智能芯片、传感器等硬件设备和云计算构成。

1. 人工智能芯片

何为人工智能芯片，至今并没有一个严格且公认的标准定义，目前通常把面向人工智能应用的芯片统称为人工智能芯片。芯片是整个人工智能体系的心脏，有“无芯片不 AI”之说。作为整个人工智能体系的根基，无论是海量数据的获取和存储，还是算法的实现和超级计算能力，都离不开芯片的强大功能。能否开发出符合市场需求、具有超高运算能力的芯片，已成为决定人工智能产业可持续发展的关键因素。

（1）人工智能芯片的特性

随着智能经济的快速崛起，市场对芯片的需求也随之迅速加大。庞大的市场规模带来了多样化的消费需求，人工智能芯片的市场由此跳出零和博弈的困局，整个行业形成明显的价值扩张，进入后摩尔定律时代，为不同功能定位的芯片提供了广阔的成长空间。

当前，能够较好满足深度学习需求的人工智能芯片是 GPU（图形处理器）。相比于适合流程控制的 CPU（中央处理器），GPU 专为并行处理而设计，广泛应用于图形和视频渲染等领域，更适合从事人工智能所需要的计算密集型任务。

AI 芯片与算法、应用的技术关联

资料来源：《人工智能芯片技术白皮书（2018）》，清华大学 2018 年版。

未来，随着人工智能算法的进步、芯片材料的多样化、芯片工艺水平提升，以及供电方法改进和能耗节约，人工智能芯片一方面会降低设计门槛，逐渐实现“平民化”；另一方面，人工智能芯片的市场竞争格局将发生改变，量子计算机和人脑模拟芯片等新型硬件的崛起，推动人工智能

芯片沿着“CPU+FPGA+GPU—特制芯片—量子芯片”的路径升级。未来的人工智能芯片需要以下四个方面的能力：大数据处理能力、数据精度、可重构能力、软件工具。

① 大数据处理能力

人工智能的发展高度依赖海量的大数据要素。满足高效能机器学习的数据处理，因此成为人工智能芯片设计中最重要考虑因素。当前人工智能芯片发展中面临一个无法回避的问题，即运算单元与内存之间的性能差距越来越大，内存子系统的孱弱性能严重阻碍了芯片整体处理能力的提高，我们将之称为“内存墙”。而人工智能的运转需要进行高负载的密集数据运算，大量的数据需要在存储硬盘与各层次存储器间频繁搬移，“内存墙”问题更加凸显。

解决运算单元与内存系统的性能差距，学术界和企业界选择了两个探索方向：一是丰富内存系统的处理单元。通过增加内存系统的存储器容量，并使其更靠近计算单元，大大减少数据计算单元和内存存储单元之间的数据移动成本（时间和功耗）。二是研发具备计算能力的新型存储器。直接在存储器内部（或更近）实现计算。这种方法也被称为存内计算（Process-in-Memory PIM）或近数据计算（Near Data Computing NDC）。[1]

---

[1] 清华大学、北京未来芯片技术高精尖创新中心：《人工智能芯片技术白皮书（2018）》，清华大学 2018 年版。

② 低精度设计

低精度设计是人工智能芯片的一个趋势，这种趋势在针对推断的芯片中尤为明显。相关研究表明，学习算法和部分神经网络更需要的是方向判断，低精度设计更符合神经形态计算的特征。多数场合使用低精度的二进制数据，就足以满足智能推断的预期效果。同时，足够用的低精度可以节省大量的内存和能耗，有利于加速机器学习的数据训练和算法推断。未来的人工智能芯片将通过分析数据上下文精度，取舍精度误差的敏感性，动态地设置和调整精度，进一步优化设计。

究竟哪些领域适合低精度设计的人工智能芯片，取决于人工智能芯片的目标应用领域。我们可以针对目标应用是“训练”还是“推断”，把人工智能芯片的目标领域分成四个象限，如下图所示。

人工智能芯片的目标领域

资料来源：清华大学、北京未来芯片技术高精尖创新中心（2018）

在上图中，云计算、大数据中心因为其高性能、高精度、大吞吐量，需要同时满足训练和推断两种任务，而训练任务必须满足高密集强度的大数据计算，必然需要高精度、可伸缩、可扩展的芯片设计。而边缘 / 嵌入设备，主要面向多种类型的不同需求，以推断应用为主，训练的需求还不是很明确。虽然部分高性能的边缘设备也需要具备一定的学习能力，存在数据训练的需求，以支持在线学习功能，但仅从硬件本身来说，它们更加类似那些部署在云端，为专门人工智能应用设计的芯片和硬件系统设备。因此，未来低精度设计的人工智能芯片将更多应用于边缘设备和嵌入式设备，这也是一个非常庞大的市场。

③ 可重构能力

人工智能还处在高速发展阶段，快速迭代是典型特征。产业快速发展的扩张阶段，各领域都提出了定制化的特色应用和算法需求，但始终面临着人力资本短缺、需求不明等众多问题。同时，定制应用和定制芯片不仅需要投入大量的人力、物力，针对特定应用和特定场景的定制化芯片和算法也很难适应外界快速迭代的周期变化，人和系统都面临着较大的机会成本。对此，未来的人工智能芯片设计，将更多考虑到芯片的研发成本和研发周期，采用针对特定领域（包括具有类似需求的多种应用）而非针对特定应用的设计原则，更加强调研发具备可重构能力人工智能芯片，

让人工智能芯片能够重新配置功能，适应新的人工智能算法、架构和任务，以满足不同定制应用的任务需求。

④ 软件工具

如同传统的 CPU 离不开编译工具的支持，人工智能芯片也同样离不开软件工具的支持。软件工具能够把神经网络和机器学习的任务要求，转换为能够在人工智能芯片上高效执行的指令代码。比如，英伟达公司的 GPU 所使用 CUDA 工具，就是一种并行计算架构的软件工具包和运算平台，能够为 GPU 提供并行计算功能，从而大幅提升 GPU 的性能，使其能够解决复杂的计算问题。长期以来，国内对于芯片、机床等硬件实体比较重视，而对工业软件、软件工具等软体相对忽视，这是国内在推动人工智能革命时，需要重点补强的环节。

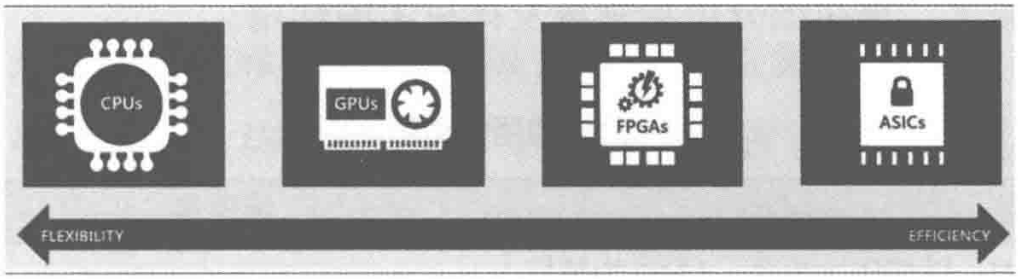
随着芯片技术的发展，软件工具被广泛应用于优化代码，其基本功能是指令代码的基本处理、内存访问以及地址分配和任务调度等。比如当软件工具链将神经网络或其他模型映射到人工智能芯片上时，可以使用神经网络剪枝、权重压缩和动态量化等技术优化代码，大幅提升人工智能芯片的运算效能。目前，人工智能算法开发框架如 TensorFlow、Caffe 和 PyTorch 等，在人工智能应用研发中已经起到了至关重要的作用。对人工智能芯片来说，构建一个集成化的流程，将人工智能模型的开发和训练，硬件

无关和硬件相关的代码优化，自动化指令翻译等功能无缝地结合在一起，将是成功部署的关键要求。

（2）云计算芯片和边缘计算芯片

① 人工智能芯片的分类

人工智能芯片根据应用环境的不同，可以分为三类：第一类是侧重加速机器学习（如神经网络、深度学习）的算法芯片。这些功能强大的芯片主要用于上游训练端，是目前计算数据中心使用的主流芯片，比如英伟达公司 1999 年发布的全球首款图片处理芯片 GPU，寒武纪公司 2016 年发布的全球首款深度学习专用处理器芯片 NPU，以及近两年以 ASIC（专用集成电路）为底芯片的谷歌 TPU、寒武纪的 MLU 等。这些人工智能芯片通常部署到云端，强调精度、处理能力、内存容量和带宽，同时追求低延时和低功耗，因此相比传统的 CPU 拥有更多的并行内核。



四类芯片的通用性与功耗平衡

第二类是各类通用人工智能芯片，主要应用于智能终端设备。因为面向产业链下游的细分市场需求，因此这些人工智能芯片往往嵌入边缘的终端设备，多进行推理判

断等应用，更加关注功耗、响应时间、体积、成本和隐私安全等问题。因此，通常会使用各种软硬件技术和工具软件，优化这些边缘设备的通用芯片，以完善功能、提升性能。目前，终端应用领域除了使用主流的 GPU 芯片之外，还存在 CPU、FPGA（可编程集成电路芯片）、ASIC（定制芯片）等多种类型的定制芯片和专用集成电路产品。比如，FPGA 具有电路级别的通用性、可编程性，能够更好地满足开发周期较短的 IoT 产品、传感器数据预处理工作以及小型开发试错升级迭代阶段等。以 TPU 为代表的 ASIC 定制芯片，针对特定算法深度优化和加速，更加适合应用需求确定的执行模型（deterministic execution model）。随着英特尔的 Nervana Engine、Wave Computing 的数据流处理单元、英伟达的 DLA 等深度学习型 ASIC 芯片面市，这些具有特定优化和效能优势的下游芯片，能够更好地满足边缘计算的需求，将在细分市场领域发挥更大的作用。

常用芯片的通用性和功耗平衡

训练端	推理端
GPU：以英伟达为主，AMD 为辅标榜通用性，多维计算及大规模并行计算架构契合深度学习的需要。在深度学习上游训练端（主要用在云计算数据中心里），GPU 是当仁不让的第一选择。	GPU：英伟达 Volta GPU 也开始布局推理端。深度学习下游推理端虽可容纳 CPU/FPGA/ASIC 等芯片，但竞争态势中英伟达依然占主导。

续表

训练端	推理端
ASIC：以谷歌的TPU、英特尔的Nervana Engine为代表，针对特定框架进行深度优化定制。但开发周期较长，通用性较低。比特币挖矿目前使用ASIC专门定制化矿机。	ASIC：下游推理端更接近终端应用，需求也更加细分，英伟达的DLA，寒武纪的NPU等逐步面市，将依靠特定优化和效能优势，未来在深度学习领域分一杯羹。
CPU：通用性强，但难以适应于人工智能时代大数据并行计算工作。	FPGA：依靠可编程性及电路级别的通用性，适用于开发周期较短的IoT产品、传感器数据预处理工作以及小型开发试错升级迭代阶段等。但较成熟的量产设备多采用ASIC。

资料来源：微软公司官网

第三类是模仿人类等生物大脑设计的神经形态计算芯片，主要用于各类脑机接口。比如，特斯拉CEO马斯克就专门成立了一家脑机接口公司Neuralink，致力于将人脑和计算机相连，通过给人脑植入芯片与计算机建立联系，可以增强人类大脑的计算和记忆能力，并治疗疾病。2023年6月，Neuralink获得了美国食品药品监督管理局的批准，得以进行人体临床试验，计划在一名四肢瘫痪或截肢的患者身上植入一个设备，以恢复患者的四肢运动功能。中国深圳的BionicM（健行仿生）公司也在高端动力假肢领域取得显著突破。但由于脑机接口问题涉及深层次的伦理道德问题，各界对此观点不一，因此本书不对此作深入讨论。

② 云计算的人工智能芯片

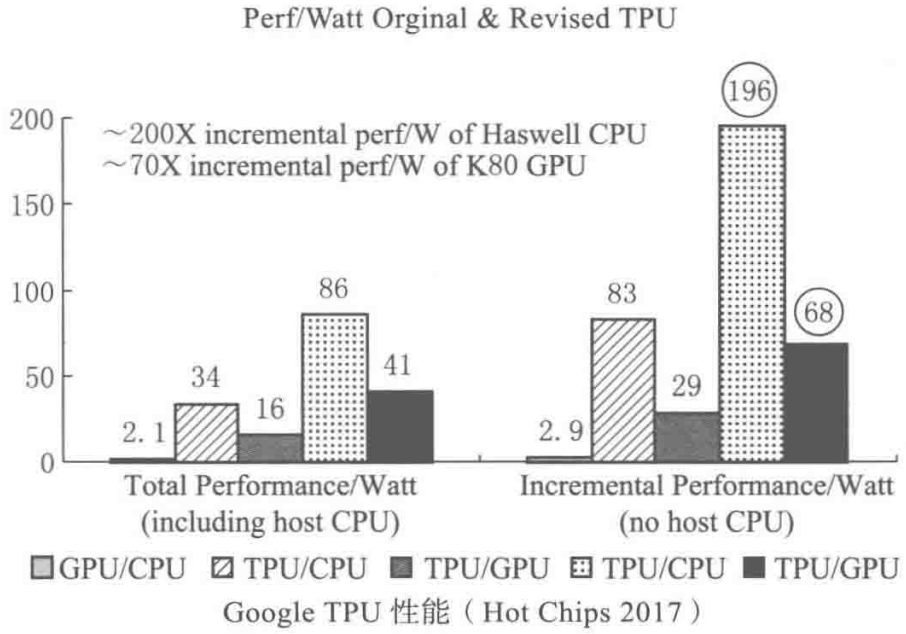
主流云计算芯片的以英伟达公司的图形处理芯片 GPU 为代表，因其优异的性能被广泛部署于云端的服务平台，用于深度神经网络的训练和推理。GPU 是一种逻辑芯片，以前被称为显卡或者图形处理器，广泛用于图像显示、图形渲染和游戏等。其与 CPU 的不同在于，一是应用领域不同。CPU 主要用于流程计算，而 GPU 主要用于并行计算。二是芯片架构不同。CPU 通常只有 1 个到数个内核，而 GPU 被称为众核，NVIDIA Fermi 有 512 个核，新一代 GPU 拥有数千个计算内核，可以实现 10—100 倍的吞吐量。三是计算方法不同。CPU 是对二进制的代码进行计算，而 GPU 是基于矢量的计算。

上述特质使得 GPU 相比 CPU 拥有更强大的计算能力。因此，原本面向视频编辑、游戏而研发的图形显卡，在人工智能时代迎来了新生，进一步升级为 GPU。GPU 为满足游戏玩家、设计师和程序员而开发的高品质视觉功能，其强大的并行运算能力受到人工智能的青睐，广泛契合监督学习以及生成式对抗网络（GAN）的多维并算需求。其通用性和灵活性非常适合处理密集数据，是强化学习所需要的。因此，GPU 成为深度学习领域的首选芯片，也顺理成章地成为云计算的主流芯片。

GPU 市场是两强领先、多家参与的格局，英伟达遥遥

领先，AMD 紧追不舍，而英特尔风光不再。深度学习领域的上游训练端由 GPU 主导，并基本为英伟达所垄断。下游推理端虽然有 CPU/FPGA/ASIC 等多种芯片，但英伟达仍然在竞争中占据主动，凭借 Volta 构架自上而下地构建了成熟的系统开发生态，成为人工智能计算领域的领导者。2023 年 4 月，英伟达发布了最新的旗舰版人工智能芯片 Tesla H100，该 GPU 采取新一代的 Hopper 架构，拥有 800 亿个晶体管，能够提供 60 TFLOPS（即每秒 60 万亿次浮点指令）的 FP64 浮点运算。英伟达公司的出色表现，不仅改写了英特尔公司制定的芯片行业规则，而且依靠相对更加完善的软件开发环境和上下游行业生态体系，成为人工智能训练领域使用最广泛的平台。

云端的人工智能应用需要强大的算力，由此造就了一



各种芯片的性能比较

个庞大的云端算力市场，仅2022年全球云端算力市场就高达3500亿美元。巨大的利润空间吸引了众多公司，纷纷开始尝试设计更高效率的专用芯片，其中尤以谷歌公司的TPU芯片最具潜力。张量处理单元（TPU）是一种定制化的ASIC芯片，使用了专用架构用于机器学习工作负载，可以支持搜索查询、翻译等应用，实现了比同时期CPU和GPU更高的效率。

著名的阿尔法狗使用的就是TPU。TPU的发展也经历不断完善，持续升级的过程。2016年谷歌公司发布了原始TPU，仅能用于推断。2018年谷歌公司又发布了第二版TPU（TPU2），在推断基础上新增了支持神经网络训练的加速功能。与同期的CPU和GPU相比，TPU可以提供15—30倍的性能提升，以及30—80倍的效率（性能/瓦特）提升，能够更好地满足搜索、街景地图、照片、翻译等业务。谷歌公司最近还开放了云服务的商用TPU，处理能力达到180TFLOP，提供64GB的高带宽内存（HBM），2400GB/s的存储带宽。

其他芯片巨头和初创公司也同样高度重视云端训练和终端推断的芯片市场。英特尔宣布推出NervanaTM神经网络处理器（NNP），该系列架构还可以优化32GB HBM2、1TB/s带宽和8Tb/s访问速度的神经网络计算。Graphcore、Cerebras、Wave Computing、寒武纪及比特大陆等初创公

司，也在积极参与云端芯片市场的竞争。

此外，FPGA 作为一种可编程芯片，一方面可以很好地支持不同的数值精度，在处理小批量任务时，具有出色的推断性能，尤其适合低精度推断，FPGA 的主要厂商如 Xilinx、英特尔都推出了专门针对人工智能应用的 FPGA 硬件（支持更高的存储带宽）和软件工具。另一方面，FPGA 支持大规模并行的硬件设计，和 GPU 相比可以降低推断的延时和功耗。因此，FPGA 逐渐在云端推断应用中占有一席之地，亚马逊、微软及阿里云等主要云服务厂商，都推出了专门的云端 FPGA 实例来支持人工智能应用。与此同时，FPGA 的可编程能力也使它可以相对更快地支持新的算法和应用，初创公司如深鉴科技等，也在开发专门支持 FPGA 的人工智能开发工具。

TPU 虽然理论上支持所有深度学习开发框架，但目前只针对 Tensor Flow 进行了深度优化。包括 TPU 在内的 ASIC 仍然面临通用性较弱、开发成本高、开发周期长等局限。即使是谷歌这样实力雄厚的大公司，面对数亿美元的开发调试费用，也不得不依靠加大 ASIC 的部署量，才能分摊设计成本。对此，各大公司采取了相应的措施，一是强化特定优化。英特尔的 Nervana Engine、Wave Computing 的数据流处理单元、英伟达的 DLA、寒武纪的 NPU 等，都是针对深度学习 ASIC。二是强化效能优势。作为专用集成

类别	GPU	ASIC	ASIC：TPU	CPU	FPGA
特点	1. 可多达上千个简单核心，上千个并行硬件线程 2. 并行运算能力、浮点运算能力强大 3. 最大化浮点运算数据吞吐量	1. 需求确定后可进行专门优化设计 2. 优秀的功耗控制 3. 性能稳定、可靠性高	1. 与 TensorFlow 深度结合，更接近 DSA（Domain-Specific-Architecture） 2. 已能同时用于高性能计算和浮点计算 3. 结合谷歌云提供云计算服务	1. 通用性强 2. 核心复杂程度高 3. 串行运算能力强，单线程性能优化 4. 晶体管空间用于复杂并行性指令（Complex ILP）	1. 电路级别的通用性 2. 可编程性 3. 适用于开发周期较短的 IoT 产品、传感器数据预处理工作以及小型开发试错升级迭代阶段
主要厂商	英伟达、AMD、Imagination 等	英特尔、德州仪器、三星、高通等	谷歌	英特尔、AMD、高通等	Xilinx、Altera（已被 Intel 收购）、Lattice、Microsemi

深度学习领域常用的芯片类型及主要芯片商

资料来源：微软 Build，谷歌官网，天风证券研究所

电路，ASIC 开发周期长，很容易面临硬件开发无法匹配软件升级的问题，以 TPU 为代表的 ASIC 定制化芯片，采取了明确细分市场需求，针对特定算法进行深度优化的策略，突出发挥确定性执行模型的应用效能，以期在深度学习领域占领一席之地。

③ 边缘计算的人工智能芯片

随着人工智能应用生态的爆发，越来越多的人工智能应用被部署到终端设备上，并在终端进行二次开发。除了主动部署的边缘计算芯片，有些场合也不得不使用边缘计算。一是采用边缘计算以提升设备的安全性。比如汽车的自动驾驶场景，由于存在网络延迟、带宽限制、用户隐私等问题，自动驾驶过程中的路况分析、车辆操控等智能推断任务，就不能交由云端完成，必须在汽车这样的边缘节点执行，以避免诸如网络延时等突发情况可能带来的灾难性后果。二是降低新型基础设施的网络负载。新型基础设施需要投入大量的人力、物力和财力，但即使再高的基础设施投入，面对海量的数据，这些基础设施依然面临着数据超载问题。比如，目前中国城市已基本普及了天网系统，大城市动辄数以百万计的高清摄像头，会拍摄海量的高清录像数据，如果原原本本地传输这些实时路况信息，将所有行人的人脸识别都交由云端处理，其数据量很可能瘫痪通信网络。因此，必须采用边缘计算。

边缘计算的人工智能芯片覆盖了一个非常广泛的范围，应用场景各式各样。自动驾驶这样的场景，需要一个性能很强的边缘计算设备，其更加关注人工智能芯片的性能，要求边缘终端设备本身必须具备足够的推断计算能力，以实现在本地就完成深度神经网络推断的需求。目前，MobileEye SOC 和 NVIDIA Drive PX 系列提供神经网络的处理能力已可以支持半自动驾驶和完全自动驾驶，能够自动处理来自多路视频摄像头、雷达、激光雷达以及超声传感器的信息输入，并融合这些数据以确定汽车的精确位置，判断汽车周围的环境，计算出安全行驶的最佳路径和相关操作。而对于可穿戴设备，则必须在非常有限的空间内，实现特定的智能功能，更加关注芯片的功耗和成本。目前边缘处理器芯片的计算能力和功耗水平都尚有欠缺，是人工智能芯片研发的重要方向。

智能手机是目前应用最为广泛的边缘计算设备，苹果、华为、高通、联发科和三星等大厂，都在研发定制的边缘计算人工智能芯片。ARM、Synopsys 等老牌 IP 厂商也在积极研发专用设备的定制人工智能芯片，这些边缘计算的 IP 产品包括手机、智能摄像头、无人机、工业和服务机器人、智能音箱及各种物联网设备。此外，地平线机器人、寒武纪、深鉴科技、元鼎音讯等初创公司也进入了边缘计算芯片和系统解决方案的领域，边缘计算的人工智能芯片将迎

来新一轮的发展热潮，FPGA 就是其中重要的一个方向。

FPGA（Field Programmable Gate Array）即现场可编程门阵列，是一种半定制的专用集成电路，属于前面所讲的 ASIC 的一种。FPGA 在 PAL、GAL、CPLD 等可编程器件的基础上进一步发展而来，采用了逻辑单元阵列 LCA（Logic Cell Array）概念，既解决了定制电路的不足，又克服了原有可编程器件门电路数有限的缺点。FPGA 在内部设置了三个部分：可配置逻辑模块 CLB（Configurable Logic Block）、输入输出模块 IOB（Input Output Block）和内部连线（Interconnect），其就是一种电路级别的结构设计，因此具有极强的通用性，与 CPU 相比更具低延迟、低功耗等优势，适用于传感器端进行数据预处理、数据等工作，是边缘计算人工智能芯片的重要发展方向，尤其适合小型开发的迭代升级使用。

需要注意的是边缘计算人工智能芯片和云计算芯片并非完全分道扬镳，云端设备和边缘设备在各种人工智能应用中往往是配合工作的，两个方向共同促进了人工智能芯片技术的快速发展。一方面，近年来，随着新型材料、工艺和器件的迅速发展，以及基础理论研究的突破，如 3D 堆叠内存、光刻工艺演进等，都显著提升了人工智能芯片的性能，大幅降低了芯片功耗。随着边缘设备能力的不断增强，越来越多的计算工作负载被安排给边缘设备执行，甚

至部分智能训练和深度学习功能，也被放到边缘设备上执行。云端计算与边缘计算分工合作的案例很多，最普遍的方式是云端负责训练神经网络，然后由边缘设备采集云端数据并进行推断。另一方面，随着边缘设备计算能力的增强，云的边界也逐渐向数据源头靠近，传统的终端设备和云端设备里将会出现更多的边缘设备，从而将人工智能的数据处理任务分散到各种网络设备中，比如现在的5G基站，就开始让数据尽量实现本地处理。顺应这种趋势，未来云端设备和边缘设备将会连同他们连接的网络，共同构成一个巨大的人工智能处理网络，它们之间的协作训练和推断将是一个有待探索的方向。

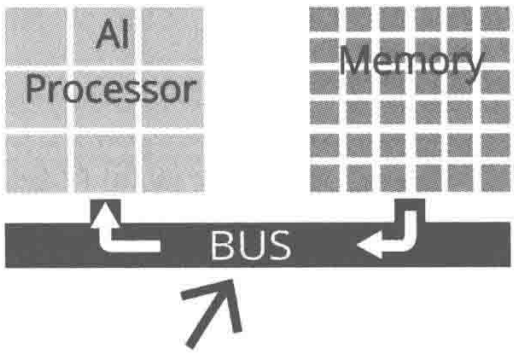
（3）人工智能芯片面临的挑战

面对巨大的应用需求，人工智能芯片一方面创新技术和工艺，提升性能，另一方面也在持续创新深度学习的计算范式。但即便如此，人工智能芯片在日益增长的高效运算需求面前，仍然面临着一些瓶颈问题，突出的有冯·诺伊曼瓶颈和CMOS工艺和器件瓶颈两种。

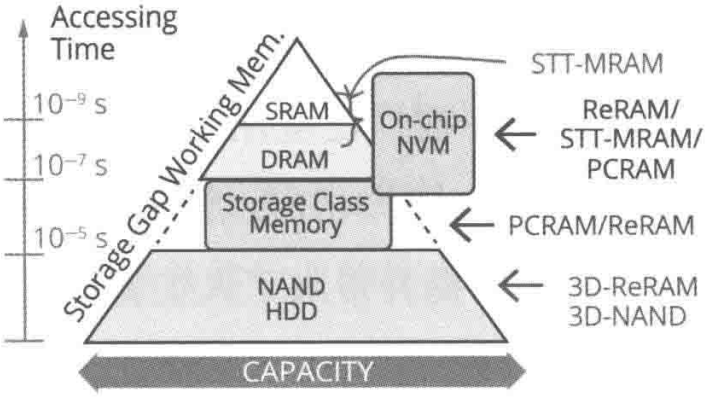
① 冯·诺伊曼瓶颈

提高人工智能芯片性能和能效的关键之一在于支持高效的数据访问，人工智能芯片的传统结构是冯·诺伊曼体系，比较简单易行。数据从处理单元外的存储器提取，处理完之后再写回存储器。由于运算部件和存储部件存在速

度差异，当运算能力达到一定程度，由于访问存储器的速度无法跟上运算部件消耗数据的速度，再增加运算部件也无法得到充分利用，即形成所谓的冯·诺伊曼“瓶颈”，或“内存墙”问题，这是长期困扰计算机体系结构的难题。



冯·诺伊曼“瓶颈”



内存层级结构

常见的解决方法是利用高速缓存（Cache）等层次化存储技术尽量缓解运算和存储的速度差异。由于人工智能芯片中需要存储和处理的数据量远远大于之前常见的应用，以常见的 VGG16 深度神经网络为例，需要 138M 个权重参数，一次推断过程需要 15.5G 次乘累加运算，这都使得人工智能时代的冯·诺伊曼“瓶颈”问题更加凸显。可以说，

目前大部分针对人工智能芯片，特别是针对加速神经网络处理而提出的硬件架构创新，都是在和冯·诺伊曼“瓶颈”问题进行博弈。

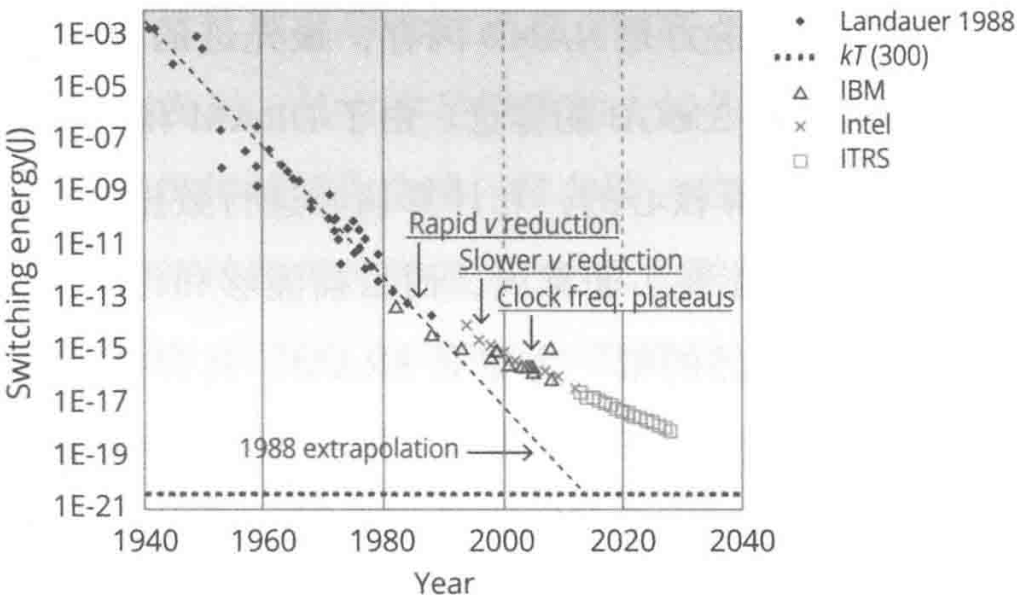
概括来说，目前在架构层面的解决思路有二种：一是减少访问存储器的数据数量，比如减少神经网络的存储需求（参数数量、数据精度、中间结果），压缩数据量，用运算性能换存储容量等。二是降低访问存储器的代价，尽量拉近存储设备和运算单元的“距离”，甚至直接在存储设备中进行运算。

② CMOS 工艺瓶颈

现今的计算机可以达到每秒千万亿次（1015 FLOPS）浮点运算的处理速度，被广泛应用于生物学、气候分析、基因组学、脑基因组学、材料开发等先进科技的研究。可以不夸张地说，算力的提升带来了人工智能的复兴，驱动着现代社会的发展。

然而，对人工智能而言，目前的芯片运算速度还是太慢了。尤其是机器学习需要更强性能的芯片，比如超过每秒百亿亿次（1018 FLOPS）运算能力的计算系统。当前，生产人工智能芯片的 CMOS 技术已基本接近理论极限，目前 CMOS 器件的横向尺寸接近 3 纳米，层厚度只有几个原子，这会显著导致电流泄漏，降低工艺尺寸缩小的效果。

不断提高系统性能主要得益于集成尺寸的缩小。然而，



逻辑器件的最小翻转能耗趋势

由于基础物理原理上限和经济的原因，持续提高集成密度将变得越来越困难，这些纳米级晶体管的能量消耗非常高，很难实现密集封装。因此，在人工智能时代，算力需求与摩尔定律之间存在鸿沟越来越大，摩尔定律接近失效，已经不能准确描绘半导体行业的技术进步。

③ 存储器件瓶颈

物联网（IoT）、社交媒体和安全设备产生了大量的数据，存储、交换和处理这些数据都需要大量的存储设备和存储单元。目前 DRAM 技术已经接近极限，提升 DRAM 性能和存储器容量，主要通过封装技术来实现，即堆叠多个 DRAM 芯片并通过硅通孔（TSV）将其连接到存储器控制器芯片上。

增加数据带宽则是通过更宽的数据总线来实现的。非

易失存储技术的主力是NAND闪存，最先进的3D NAND具有多达64层和256GB的容量。由于DRAM和NAND闪存都是独立于计算核心的，和计算核心进行数据交换的代价（包括时间和能耗）非常大，而目前能够和计算核心紧耦合的片上存储器的唯一方案是SRAM，其容量为兆级。即使采用最小的SRAM单元填充1平方厘米芯片面积的一半，也只有约128兆的片上存储容量。因此，发展人工智能需要解决数据存储的瓶颈，尤其是需要在智能芯片的架构层级，去解决存储和运算的矛盾。

2. 超级计算机

人工智能的大数据模型需要对海量数据进行高性能计算，这种高负载的任务通常只有使用超级计算机才能完成，因此超级计算已经成为人工智能的代名词。所谓超级计算机是指具有极大的数据存储容量和极快速的数据处理速度的超大型电子计算机，其基本组成组件与个人计算机基本相同，但规模与性能则强大许多。以每秒浮点运算次数（FLOPS）衡量超级计算机的性能，速度为1 petaflop的计算机系统每秒可以执行百万的四次方浮点运算（1015 FLOPS），处理能力比最快的笔记本电脑高出一百万倍。因其超强的性能，超级计算机可以在多种领域执行一般PC机无法处理的海量数据与高速运算。

高性能超级计算机为解决工程和科学中的重大难题而

生，是世界发达国家争抢的重要“制高点”，对国家安全、经济和社会发展，具有举足轻重的支持作用。目前，美国在超级计算机领域领先全球，排名第一的是IBM公司和美国能源部橡树岭国家实验室（ORNL）联合研发的Summit超级计算机，其拥有2397824个处理器，峰值速度达到200795 TFlop/s，比中国性能最强的神威·太湖之光超级计算机快了近60%。排名世界第三的是中国自主研发的“神威·太湖之光”超级计算机，其处理器为众核CPU“申威26010”，整台“神威·太湖之光”拥有10649600个处理器，峰值速度为125436 TFlop/s。

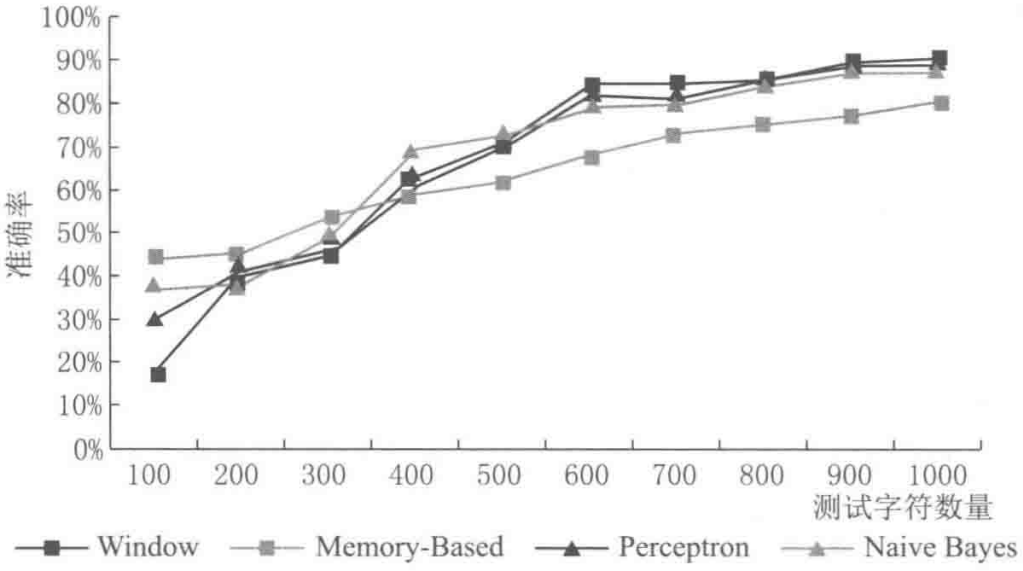
超级计算机就是人工智能领域的标杆和旗帜。超算领域的领导地位与国家安全直接关联，只有通过超级计算机的研发，才有机会成为全球人工智能研究中心，带动人工智能上下游产业链的进一步发展。目前，中国有6个超级计算中心，已经构建了从编译器、操作系统、开发环境到应用平台的较为完善的软件生态链。应用领域涉及天气气候、航空航天、先进制造、新材料等众多方面，尤其在工程技术领域，意义重大。中国在超算领域虽然已经进入全球第一阵营，但相比美国仍然相对落后。

未来突破的领域，一是真正用于人工智能的超级计算机芯片还只是处于CPU、GPU层面，如何将更适用于网络神经算法的NPU、FPGA等芯片量产化并融入超级计算机芯片

矩阵，是在人工智能发展的第一阶段，即技术驱动阶段应该重点努力的方向之一。二是超级计算机能被用在什么领域，是未来竞争的关键点。超级计算机在航空航天、先进制造、生物医药、新材料、新能源、气候模拟、地震模拟、工业仿真等领域的广泛应用，以及为人工智能领域的深度学习服务，都能带动人工智能相关上下游产业链的创新发展。

3. 海量数据

海量数据能够为算法实现和优化提供有力支持，与深度学习和训练的准确性密切相关。研究显示，数据字符数越多，训练得越充分，算法模型结果的准确率越高。



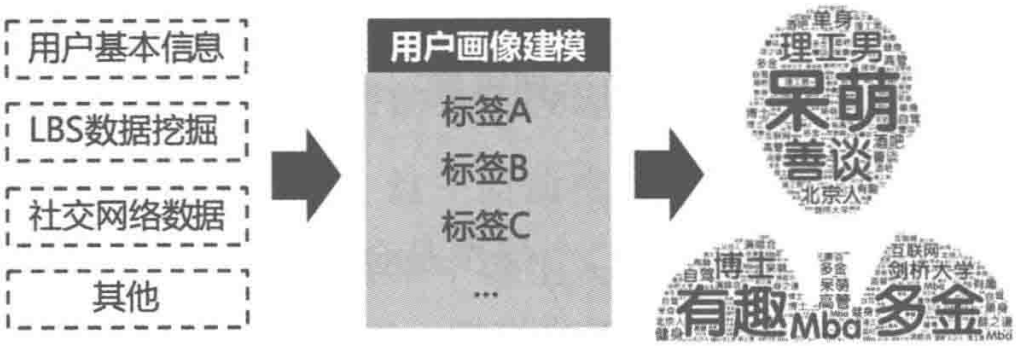
说明：window、mermory-based、perception、naive bayes 均为不同算法

数据输入量及算法模型准确性的关系

资料来源：Stanford 机器学习公开课，36 氪研究院

海量数据对人工智能革命的影响，主要体现在人工智能发展的第二阶段。此时算法与算力的矛盾已经缓解，不

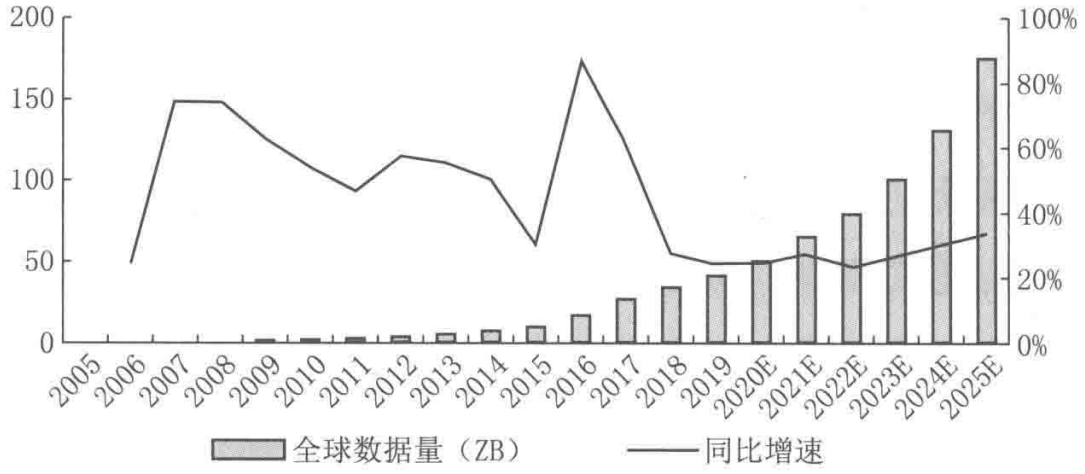
可逾越的壁垒基本消除，技术创新让位于产业化的速度竞争，数据成为基础的生产要素，驱动人工智能向细分行业、虚拟场景进行融合应用，并持续迭代更新。在此阶段，大量结构化、可靠的数据被采集、清洗和积累，甚至变现。例如，在海量数据的基础上，可以精确地描绘出消费者的用户画像，以此为依据制定出的消费者个性化营销方案，能够大幅提高产品销售的成单率，缩短交易时间和交易成本，从而提升整个社会的运行效率。



从数据到用户画像
资料来源：36 氪研究院（2017）

随着移动互联网、电子商务等领域的快速发展，全球数据量实现了爆发式增长，这为人工智能的发展奠定了坚实的基础，这是人工智能技术和产业能够快速崛起的重要原因。

根据 IDC、希捷的联合研究，人类已经进入了数字经济时代，未来较长时间内，全球的数据量都将迎来惊人的增长，2025 年全球的数据量将达到 163ZB，是 2016 年的 10 倍之多，其中 2015—2025 年的平均增速将超过 30%。如



2005—2025 年全球数据量（含预测）

资料来源：IDC、平安证券研究所、中信证券研究所

此海量的数据既为人工智能的发展创造了条件，也带来了如何应对海量数据冲击的治理难题。

（三）技术层

人工智能革命的第二个层次是金字塔框架图中的技术层，这一层次主要提供算法模型，以及相关的感知功能和认知功能。在有的分类方法中，把技术层拆分为算法模型的算法层和实现各类感知功能的感知层，考虑到人工智能的感知和认知都是基于算法模型实现的，本质上都是基于算法模型的通用技术，故而在本书中，我们将这两个层次合而为一，统称为人工智能的核心技术层。

各类感知和认知功能是人工智能的外在表象，为了实现这些拟人化的功能，人类发展了语音识别、图像识别、机器学习、深度学习等各类人工智能技术，开发了众多的算法模型，并通过海量识别训练和机器学习建模，持续完

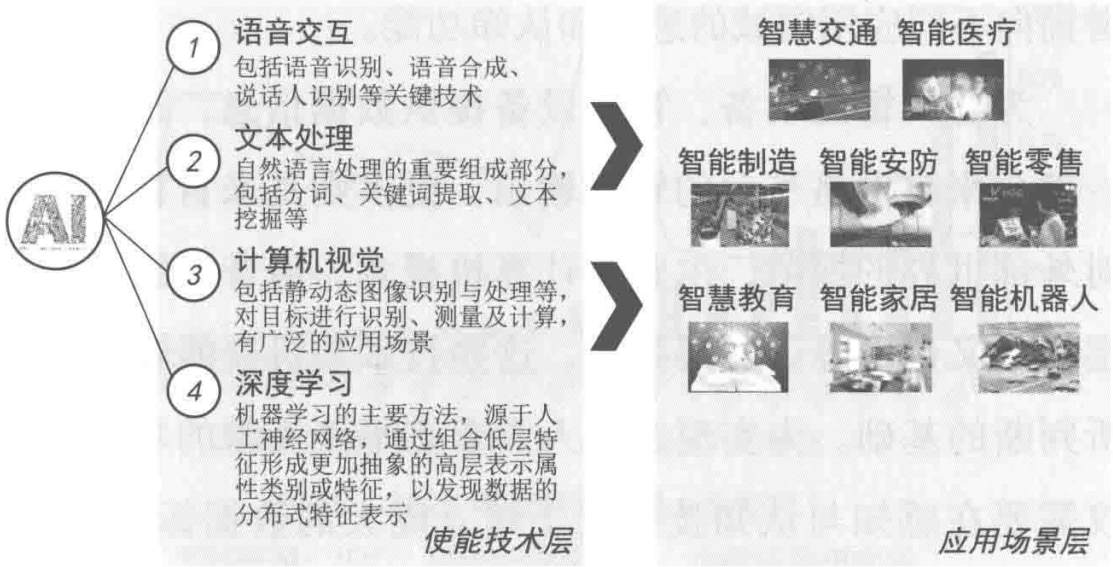
善面向不同应用领域的感知和认知功能。

为各类智能装备、智能设备提供数据信息，这些信息一部分来自网络平台的输入输出，更多则是来自智能单元对外部世界的探测，主要由计算机视觉、语音识别等感知层和语义识别等认知层构成，这些技术是机器能够做出分析判断的基础。为实现这些人类看来平淡无常的功能，不仅需要在感知与认知技术之下建立庞大的数据标注体系，为人工智能模拟现实世界提供映射和连接，而且需要机器学习、深度学习等一系列底层技术提供算法支撑。

根据人工智能发展的三阶段理论，人工智能革命目前仍处于技术驱动阶段，这一阶段推动人工智能发展的主要力量是算法和算力，会集中诞生人工智能的基础理论、基本规则和基础开发工具。在这些底层的算法支撑技术中，主流应用是基于多层神经网络的深度算法，一方面不断加强从海量数据库中自动归纳出物体特征的能力，另一方面不断加强对新事物的多层特征提取、描述和还原能力。语音交互、文本处理、计算机视觉和深度学习是当前人工智能革命最为核心的前沿细分技术。

1. 核心技术进展

业界通常把“语音交互、文本识别、计算机视觉和深度学习”视为人工智能的四项核心通用技术。其之间的关联关系源自近年来以深度学习为代表的人工智能技术飞速



人工智能的核心技术与重点应用场景

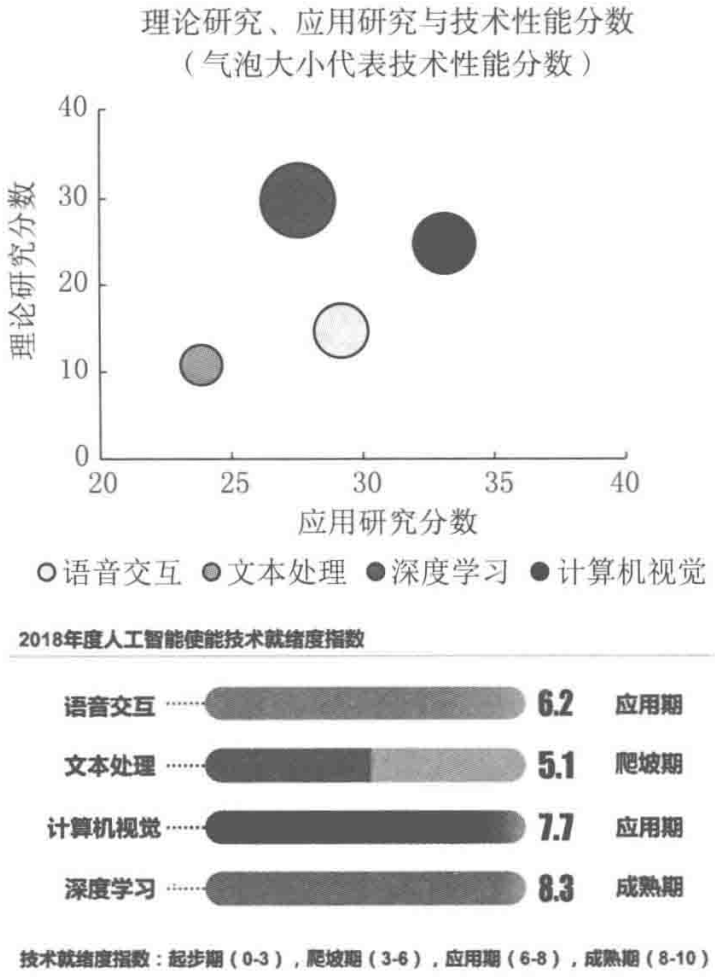
发展，引起了计算机视觉、自然语言处理、语音识别等领域的广泛关注，并成为深度学习技术应用的热点领域，取得了众多突破性的进展和应用成果，推动了各类 AI+ 项目的落地应用。

四项核心通用技术可以从理论研究、应用研究和技术性能三个维度进行评估。其中，在理论研究维度主要有 6 项参考因素：会议论文数量、会议论文被引频次、期刊论文数量、期刊论文被引频次、高被引论文数量、研究机构数量。在应用研究维度主要因素也有 6 项：专利数量、平均专利强度、PCT 国际申请比例、机构数量、人才数量、科技企业与研究机构比例。在技术性能维度主要有 1 项指标：技术性能表现。

综合理论研究、应用研究和技术性能三个维度，从四项使能技术的就绪度指数可以看出，深度学习是理论研究

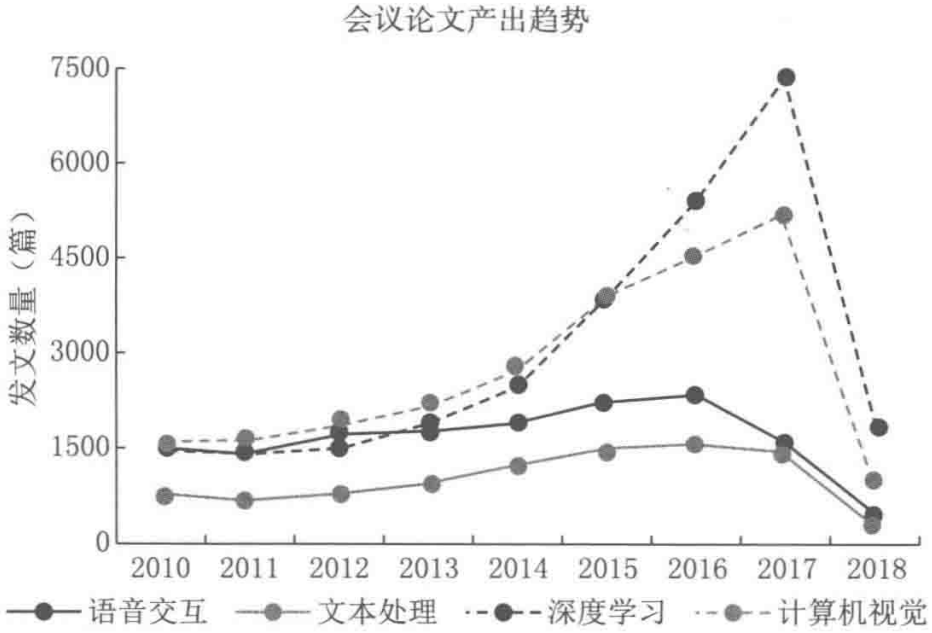
的重点方向，而计算机视觉是应用研究的重点方向。当前，深度学习的发展最为迅速，技术就绪度最高（8.3），处于技术成熟期；文本处理的就绪度最低（5.1），仍处于技术爬坡期；语音交互和计算机视觉处在技术应用初期。

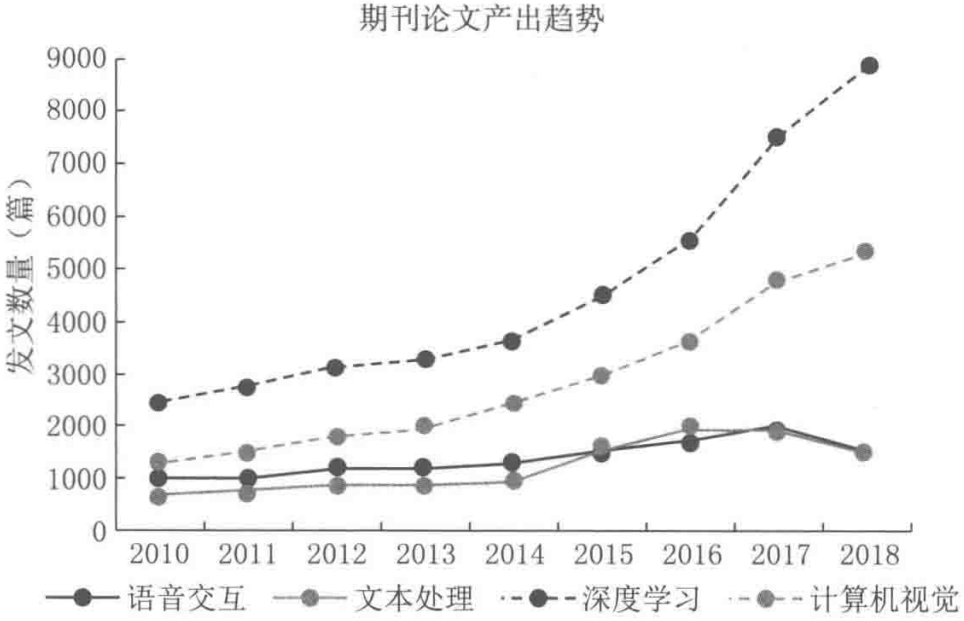
从论文产出趋势来看，人工智能的论文从2013年开始呈现快速增长趋势，其中深度学习技术的论文产出增速最高，而语音交互和文本处理的产出基本保持平稳。从论文产出规模来看，人工智能的基础研究主要集中在深度学习

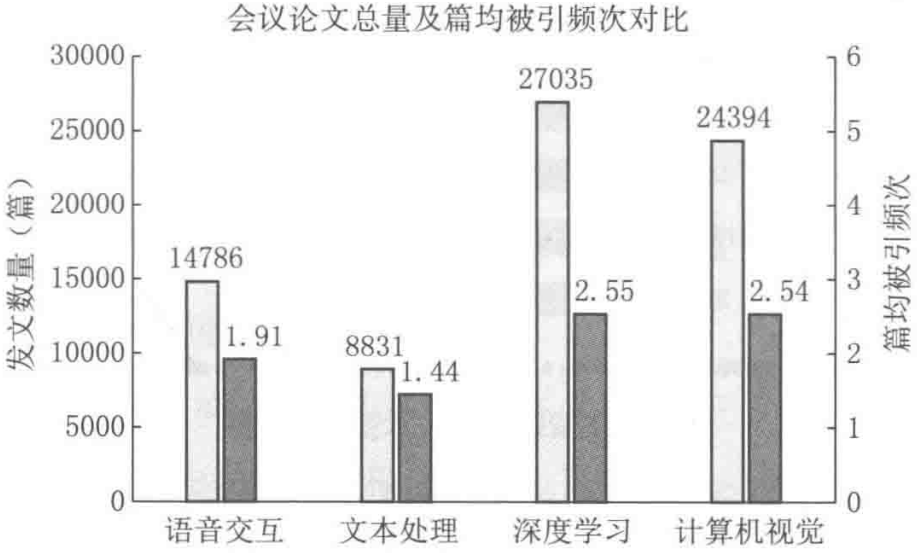


人工智能使能技术就绪度指数

资料来源：华夏产业研究院（2018）

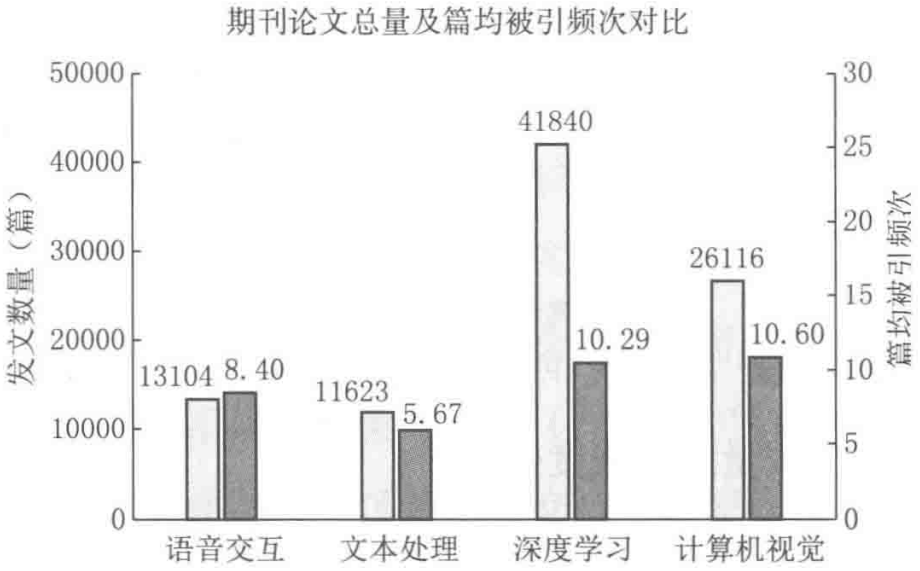

会议论文产出趋势
发文数量（篇）
7500
6000
4500
3000
1500
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
语音交互
文本处理
深度学习
计算机视觉


期刊论文产出趋势
发文数量（篇）
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
2010
2011
2012
2013
2014
2015
2016
2017
2018
语音交互
文本处理
深度学习
计算机视觉


会议论文总量及篇均被引频次对比
发文数量（篇）
篇均被引频次
30000
25000
20000
15000
10000
5000
0
6
5
4
3
2
1
0
14786
1.91
8831
1.44
27035
2.55
24394
2.54
语音交互
文本处理
深度学习
计算机视觉

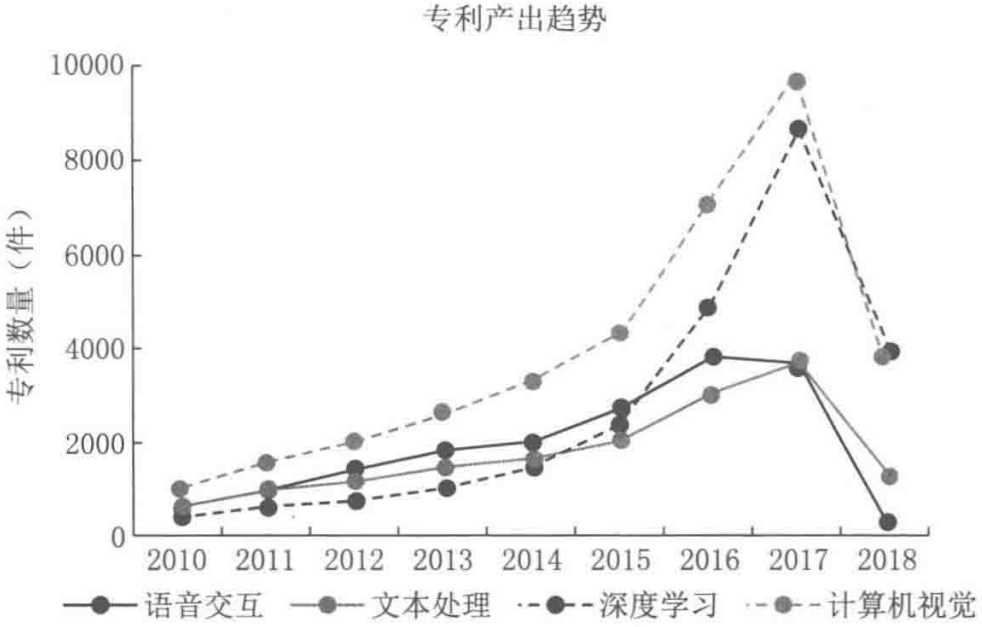
和计算机视觉，而文本处理方向的论文产出最少，份额仅为 12%。从论文的影响力来看，计算机视觉和深度学习论文引用频次领先，而文本处理技术的论文引用频次较低。

从四项核心技术的应用研究看，深度学习和计算机视



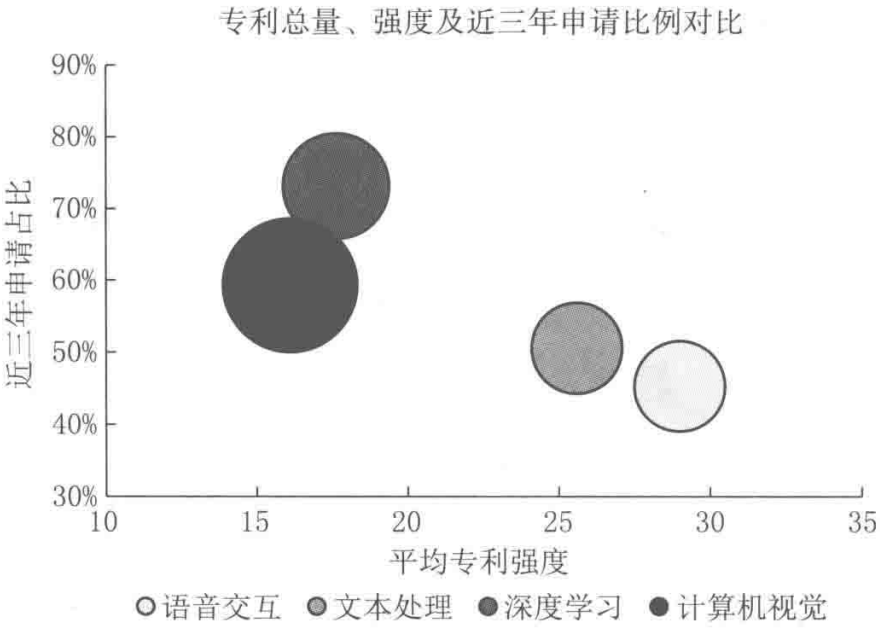
人工智能的论文成果

资料来源：Web of Science



人工智能专利产出趋势（2018）

资料来源：INNOGRAPHY（检索时间截至 2018 年 10 月）



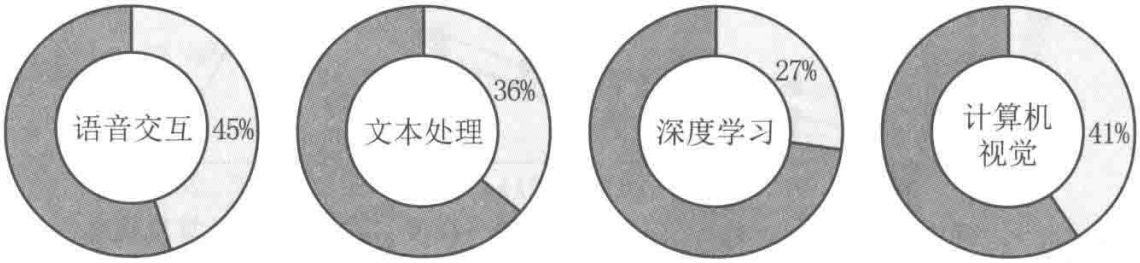
人工智能专利总量和强度（2018）

资料来源：INNOGRAPHY（检索时间截至 2018 年 10 月）

觉专利申请数量多但强度低。

从专利申请趋势来看，深度学习和计算机视觉的专利申请从 2015 年开始快速增长，语音交互的专利申请自 2016 年开始保持稳定。

综合近三年专利申请比例和平均强度来看，语音交互专利申请比例低但平均强度较高，表明语音交互已不再是近期关注重点，计算机视觉和深度学习专利申请占比较高，但平均专利强度较低，专利布局仍处于起步阶段。

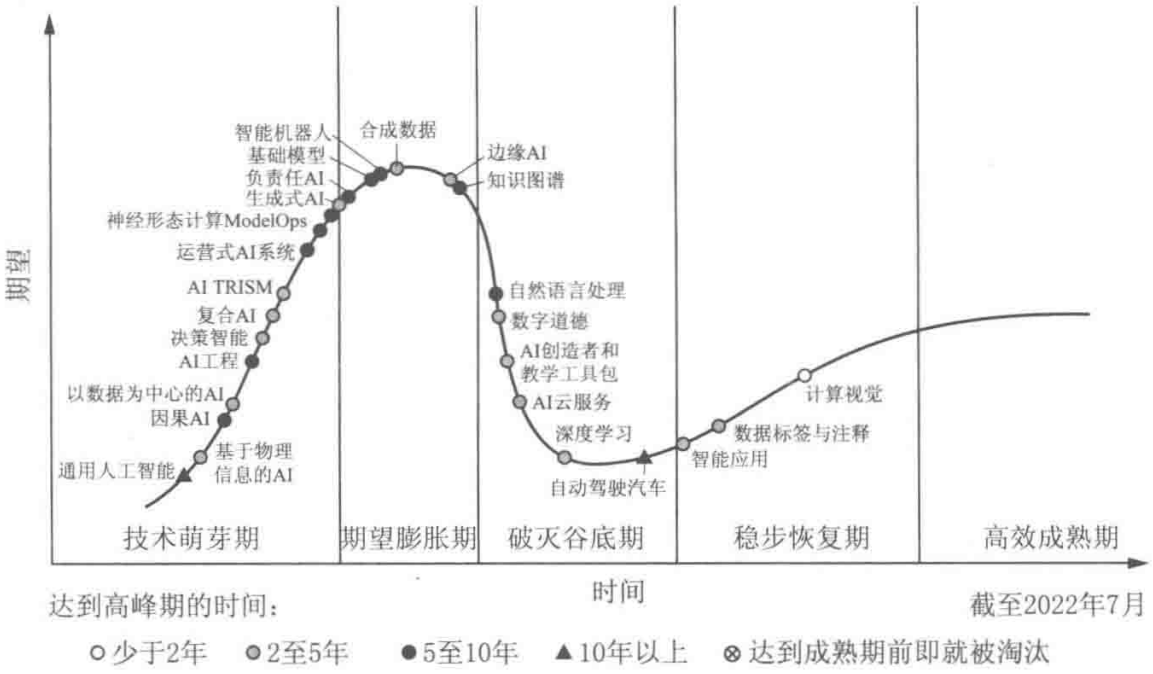


使能技术专利授权比例

资料来源：华夏产业研究院（2018）

从专利授权比例来看，语音交互技术的授权率最高（45%），深度学习技术的授权率最低（27%），主要是因为深度学习的大量专利申请仍在审查期，导致授权比例偏低。

展望未来，人工智能还在发展更多的核心技术，一个多核心的人工智能技术网络正在形成。根据 Gartner 公司最新发布的《2022 年人工智能技术成熟度曲线》报告，其将人工智能技术划分为输入端、中间体和输出端三大主体。在输入端，计算机视觉技术已经步入了成熟期，在企业和社会都得到了广泛应用，应用场景丰富，工程实践众多。而对于自然语言处理（NLP）的期望则在缓慢下降（见下图“破灭谷底期”的最上方），计算机与人类的智能化交流依然面临着可靠性与可信度的问题，语音交互、文本处理等更多被当成人类的辅



2022 年人工智能技术成熟度曲线

资料来源：Gartner 官网

助工具和参考提示，主要应用于“人工智能增强”的场景中。

根据“人工智能技术成熟度”曲线，以数据为中心的人工智能、基于物理信息的人工智能和因果人工智能都处于技术萌芽期，三者试图以不同的技术路线突破人类的认知障碍和认知边界。传统的机器学习（ML）方法是根据统计关系（相关性）进行预测的，而不考虑这些统计关系之间是否真正存在因果关系。比如，机器学习方法中的深度学习模型就像一个黑箱，数据输入后人们很难解释其输出结果的运行过程，因而也很难确定这些输出结果是否可靠，是否真正存在因果关系。因此，复合人工智能技术应运而生，试图将基于“连接主义”的数据驱动型人工智能，与基于规则推理、图型分析、主体建模和优化的“符号型人工智能”进行融合，以机器学习、数据驱动的方式，结合图技术、优化算法等等，让抽象化机制发挥更大作用，从而减少大数据模型所需要的数据和能量。

这种复合人工智能的技术思路，其典型代表就是自动驾驶技术（见上图中间“破灭谷底期”的底部）。以特斯拉的自动驾驶技术为例，其采取了纯视觉的技术路线，以计算机视觉技术进行车辆识别、行人识别，深度学习在其中发挥了非常重要的作用。但同时，自动驾驶技术也融合了更多其他的人工智能技术，比如特斯拉自研的神经网络系统，可以模拟人类的驾驶习惯，即用眼睛看、用大脑思考、

用自身的经验和实际情况等来做出加减速、踩刹车等动作，这都离不开以大数据为中心的模型和基于物理信息的人工智能系统。相信随着科技的飞速发展，越来越多的人工智能技术会进入成熟期，从生产方式、运营模式、产品服务等众多方面为各行各业赋能，给企业和社会创造更大的价值。

2. 机器学习算法

人工智能的使能技术发展迅速，机器学习和深度学习已经成为当前主流的人工智能算法，是理论研究的重点方向。作为人工智能的核心，算法的实质是一个学习的过程，其中关键是解决“学什么”“怎么学”“做什么”三个问题。“学什么”需要找到能够表征需求的模型，明确输入、输出之间的映射关系。“怎么学”是缩小函数模型结果与真实结果误差的过程，需要通过持续的迭代优化，最终将误差锁定在可接受的范围内。“做什么”则有分类、回归以及聚类三项任务。

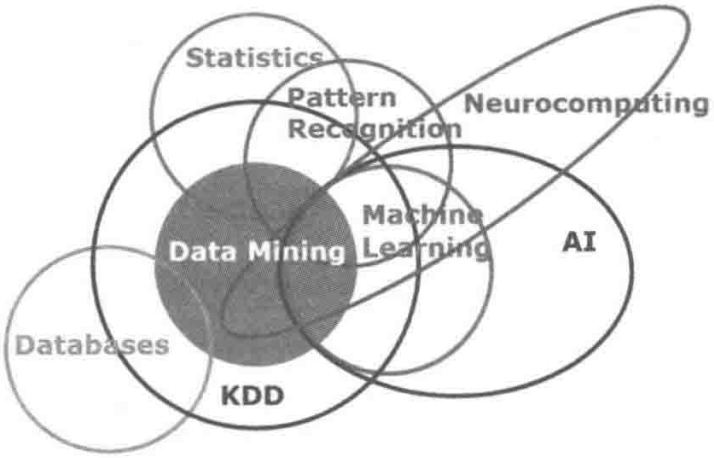
（1）机器学习的概念

机器学习（Machine Learning）涉及的领域非常广泛，由于发展迅速，变化也相当明显，因此，至今对“机器学习”并没有一个简要的定义。追根溯源，图灵早在 1950 年就提出了机器学习的概念。1952 年，“机器学习之父”亚瑟·塞缪尔（Arthur Samuel）设计了一款可以学习的跳棋

程序。这个程序构建了一个观察棋子走位的学习模型，可以用来提高下棋技巧。通过与跳棋程序的多场对弈，塞缪尔发现程序的棋艺会随着对弈次数的增多而变强，由此在1956年正式提出了机器学习的概念，认为机器学习是在不直接针对问题进行编程的情况下，赋予计算机学习能力的一个研究领域。

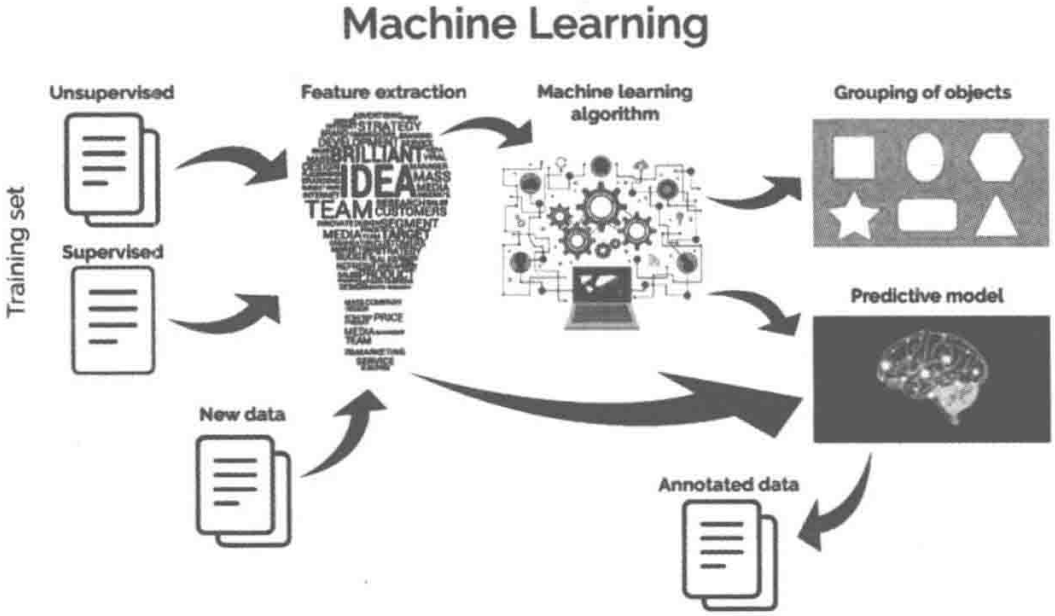
“全球机器学习教父”汤姆·米切尔（Tom Mitchell）在《机器学习》一书中，将机器学习定义为：对于某类任务T和性能度量P，如果计算机程序在T上以P衡量的性能随着经验E自我完善，就称这个计算机程序从经验E学习。这一定义过于学术和抽象，我们简而言之地认为机器学习是人工智能的一个重要子领域，其算法和处理系统的特色是通过找出数据里隐藏的模式，以此做出目标识别和预测判断。其过程可以表述为“数据预处理—特征处理（特征提取、转换和选择）—机器算法学习、训练（核心过程）—预测”。

由于机器学习的内涵和外延都在不断地变化，机器学习与数据挖掘（Data Mining，DM）和知识发现（Knowledge Discovery in Database，KDD）存在较大的领域交叉，人工智能、机器学习、数据挖掘、模式识别（Pattern Recognition）、统计、神经计算（Neuro Computing）、数据库、知识发现（KDD）等概念的交叉关系图如下：



机器学习相关概念的辨识

总体而言，机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科，是一门专门研究计算机怎样模拟和实现人类学习行为的技术。其作用机理是通过获取新的知识或技能，重新组织已有的知识结构，不断改善自身的性能。其学习过程见下图：



机器学习基本过程

资料来源：《中国工程院知识智能联合研究中心——人工智能发展报告》，清华大学 2019 年

（2）机器学习的经典算法

机器学习根据算法不同可以分为多种类型。按函数不同，机器学习有线性模型和非线性模型两种算法。按照学习准则不同，机器学习有统计方法和非统计方法两种算法。我们在日常工作中接触的是产业界的分类方法，是按照训练样本提供的信息以及反馈方式的不同进行分类的，主要有四种：监督学习（Supervised Learning）、无监督学习（Unsupervised Learning，UL）、半监督学习（Semi-Supervised Learning）、强化学习（Reinforcement Learning，RL）。

① 监督学习

监督学习使用的数据集是有标签的，也就是说我们事先知道研究数据的样本答案。根据数据标签的类型不同，监督学习可以分为分类问题和回归问题两类。分类问题是根据离散数据，预测某件事物的类别，比如给定一个人的年龄、身高、体重等信息，然后判断其性别、健康状况等其他信息。回归问题是根据连续的数据，预测某一样本所对应的数据输出，比如根据某一地区人群历年来的平均身高数据，预测该地区人群未来的平均身高。我们在工作中接触到的大部分模型都是监督学习，包括线性分类器、支持向量机等。常见的监督学习算法有：k- 近邻算法（k-Nearest Neighbors，kNN）、决策树（Decision Trees）、朴素贝叶斯（Naive Bayesian）等。

② 无监督学习

相比监督学习，无监督学习模型中没有监督者（Supervisor）。因此，无监督学习中的数据集是完全没有标签的。其对样本分类时，先假设相似样本存在地理临近、空间临近的特征，据此可以采取关联分析、聚类问题和维度约减等方法，对数据进行分类。

关联分析是指发现不同事物之间同时出现的概率。关联分析在购物篮分析中被广泛使用。比如，人工智能算法发现买面包的客户中，有 80% 的概率会购买鸡蛋，那么商家就会根据人工智能提供的信息，把鸡蛋和面包放在相邻的货架上。人们在刷短视频时，之所以总会刷到自己感兴趣的视频，也是同样的道理。

聚类问题是指将相似的样本划分为一个簇（cluster）。与分类问题不同，聚类分析预先并不知道问题的类别，自然训练数据也没有类别的标签，而是通过算法模型将数据分为不同类别。

维度约减是指减少数据维度的同时，保证不丢失有意义的信息。维度约减通常利用特征提取和特征选择的方法，达到维度约减的效果。特征选择是指选择原始变量的子集，通常有过滤法（Filter）、包裹法（Wrapper）和嵌入法（Embedded）三种方法。特征提取是将数据从高维度转换到低维度，比如主成分分析就是一种特征提取的算法。

③ 半监督学习

半监督学习（Semi-Supervised Learning）是监督学习与无监督学习相结合的一种学习方法。半监督学习一般针对的数据样本量较大的问题，通常用于解决标签数据较少，或者标签数据的难以获取的情况。与使用所有标签数据的监督学习模型相比，使用训练集的半监督学习模型在训练时可以更为准确，而且训练成本更低。

常见的半监督学习有两种：直推学习（Transductive learning）和归纳学习（Inductive learning）。其中直推学习基于“封闭世界”的假设，模型在学习阶段除了能够看到训练数据（样本 + 标签）外，还能够接触到测试数据（样本），在学习的过程中模型就已经知道未来要预测的示例是什么样的。这种学习假设，人们通常并不关心决策函数在整个样本空间会有怎样的性能表现，而只是期望在给定的预测样本上，能够达到最好的性能。因此，直推学习模型在学习过程中，会把测试数据作为一个整体去预测其标签，而不是孤立地、一个个地预测每一个测试样本。

归纳学习依赖经验数据，通过总结数据间的相似性，以发现规律，因此又称为经验学习、相似性学习。归纳学习基于“开放世界”假设，即模型进行学习的时候不知道未来要预测的示例是什么，通常被用于符号学习，是符号学习中应用最广的一种方法。归纳学习的过程是根据某个给定概

念的经验数据，从大量正例和反例中，归纳推导出一般的判定规则和模式，然后去预测未知的样本。并获得新的概念，创立新的规则，发现新的理论。常见的逻辑回归、GBDT（集成算法的一种）、神经网络等都是典型的归纳学习模型。

④ 强化学习

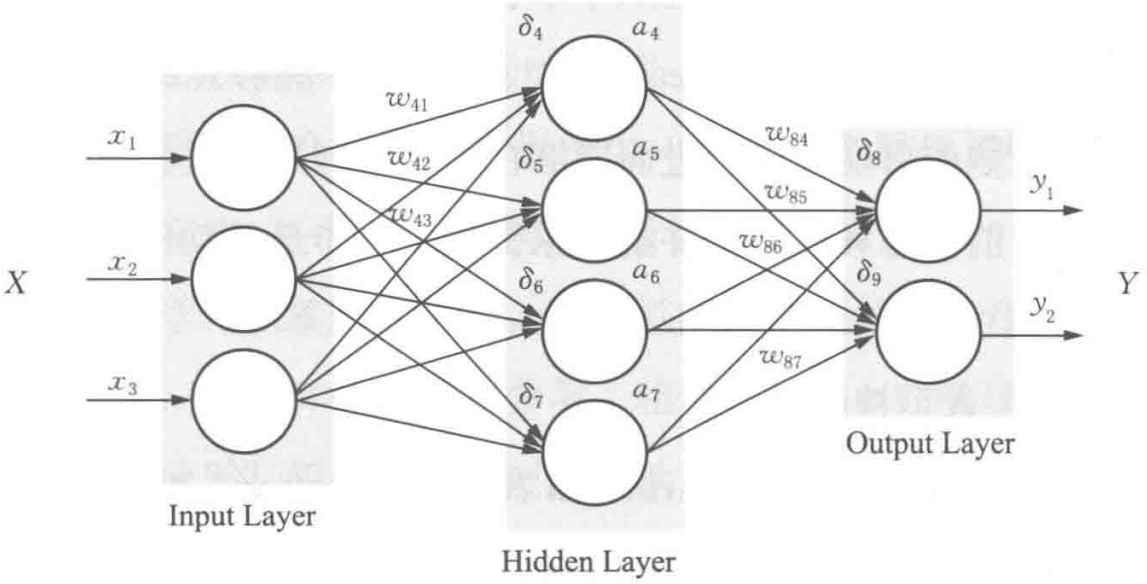
强化学习（Reinforcement Learning，RL）从动物学习、参数扰动等自适应控制理论发展而来，基本原理是：如果模型（Agent）的某个行为策略导致环境的正奖赏（强化信号），那么 Agent 以后会强化这个行为策略。Agent 的目标是在每个离散状态下发现最优策略，以最大化期望的奖赏。比如，在扫地机器人的训练过程中，扫地机器人与障碍物发生碰撞后，传感器会给机器人发送负面的反馈信息，机器人总结这些反馈信息后，会从中学会如何避免同样的碰撞。又如，在视频游戏中，Agent 通过反复试验，能够总结出获得更高分数的可能动作，进而能够在游戏中利用回报去理解玩家的目前状态和喜好，并给玩家提供下一步最佳动作的相关建议。

⑤ 人工神经网络

人工神经网络（Artificial Neural Network，ANN）是由大量处理单元互联组成的非线性、自适应信息处理系统，具有自我学习、联想存储、高速寻找最优解等功能。人工

神经网络是一种模仿动物神经网络的行为特征，进行分布式、并行信息处理的算法模型。其基本过程可以概述如下：外部刺激通过神经末梢，转化为电信号，传导到神经细胞（又叫神经元）；无数神经元构成神经中枢；神经中枢综合各种信号，做出判断；人体根据神经中枢的指令，对外部刺激做出反应。

人工神经网络经历了漫长的发展阶段。最早是20世纪60年代提出的“人造神经元”模型，叫做“感知器”（perceptron）。随着反向传播算法、最大池化（max-pooling）等技术的发明，神经网络进入了飞速发展的阶段。神经网络就是将许多个单一“神经元”联结在一起，一个“神经元”的输出就可以是另一个“神经元”的输入。一个典型的人工神经网络结构如下图所示：



典型的人工神经网络结构

（3）机器学习的未来

目前，机器学习已经进入全面爆发期，增强学习、自动机器学习、联邦学习、深度学习、可持续发展机器学习都是未来的发展方向，具体朝哪个方向发展，未来主要取决于产业化进程中的业务需求。回顾机器学习的发展历程，杰弗里·辛顿（Geoffrey Hinton）于2006年在《科学》杂志（*Science Magazine*）上发表了一篇文章，提出了神经网络的Deep Learning算法，大大提高了神经网络的性能，开启了深度学习的浪潮。2015年，杨立昆（Yann LeCun）、约书亚·本吉奥（Yoshua Bengio）和杰弗里·辛顿（Geoffrey Hinton）在机器学习的基础上，联合综述了深度学习算法，用多个处理层的计算模型，来学习具有多层次抽象的数据表示。比如，要学习到某种语义表示，通常机器学习模型需要从底层特征分析入手，经过多步非线性转换，才能得到底层特征，再以此为基础，总结出中层特征、高层特征，这就是一个深度学习的模型。

与机器学习相比，深度学习模型的优势在于，其能够从数据中自动学习到有效的特征表示，并通过多步转换为更高层次、更抽象的表示，从而代替机器学习过程中的人工干预，省去了耗费惊人的人力“特征工程”。深度学习的出现，大大提升了机器学习的效率，让图像、语音等感知类问题取得了真正意义上的突破，进入了实用化的阶段，

人工智能也因此进入了全新时代。

3. 深度学习算法

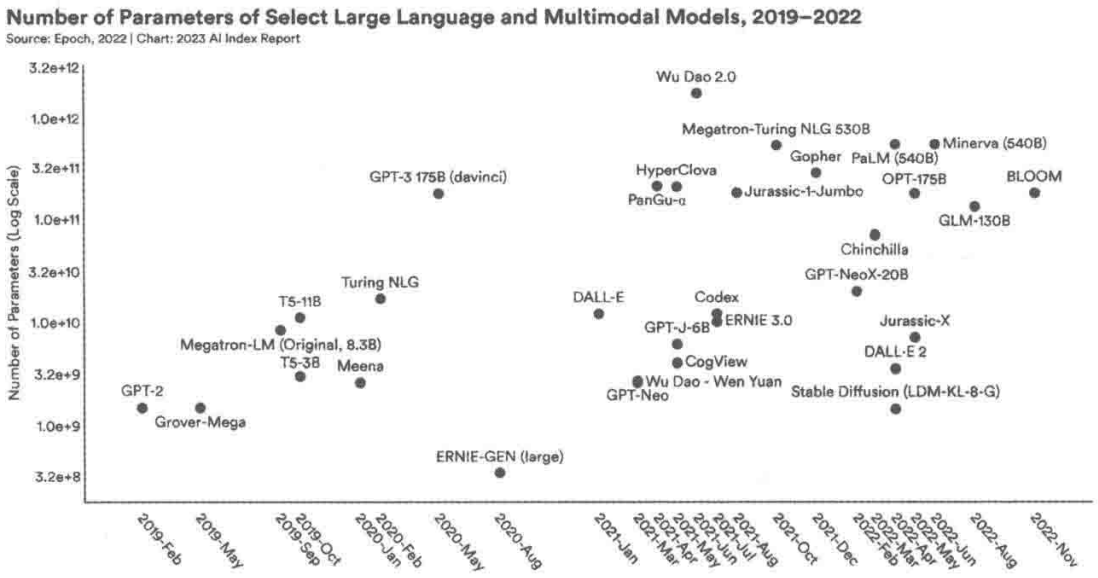
（1）深度学习的崛起

① 算力提升与深度学习

深度学习的概念早在20世纪60年代就已经提出，之所以近十年来才成为机器学习领域发展最快的分支，并成为强人工智能的核心技术，主要还是得益于摩尔定律的突破，以及并行计算和云计算的飞速发展，使得深度学习模型跟随人工智能得以普及。长期以来，计算能力严重制约了人工智能的发展，表现为：随着人工智能模型日益大型化、复杂化，智能模型的计算参数和计算量也在快速增长，对浮点运算的性能要求更是呈指数级的增长，传统的CPU和计算机已经无法满足这种快速增长的运算需求，导致人工智能陷入低潮。而数据训练的匮乏，让深度学习的效率甚至还不如人类工程算法（human engineering algorithm）。所以，长期以来人们普遍认为深度学习虽然很有趣，但效率不高、用处不大。直到近年来大数据、大模型、GPU共同取得突破，算力提升才使得深度学习/神经网络获得惊人发展。

比如，在计算能力方面，2000年英特尔推出了奔腾CPU，能够支持微软的Windows XP系统，奔腾的速度仅MIPS（百万个指令/秒），而2020年人工智能芯片已达ExaFLOPS级（百亿亿次浮点运算），运算速度是奔腾芯

片的数千亿倍。在模型参数方面，2015 年微软 ResNet 模型只含有 6000 万个参数，运算量为 7ExaFLOPS。2016 年百度语音识别系统 Deep Speech 2 的参数量已上升到 3 亿个，运算量提升至 20 ExaFLOPS。而 2017 年谷歌公司推出的 NMT 神经网络机器翻译系统，参数量达 87 亿个，运算量达到 105ExaFLOPS。2022 年 ChatGPT-3 模型的参数已达约 1750 亿个，预训练数据量为 45TB，折合成训练集约为 3000 亿 tokens。按照有效算力比率 21.3% 来计算，训练阶段实际算力需求为 1.48 × 109 PFLOPS。如此海量的人工智能模型参数和庞大的数据计算量，在 2000 年的奔腾时代，显然是无法实现的。



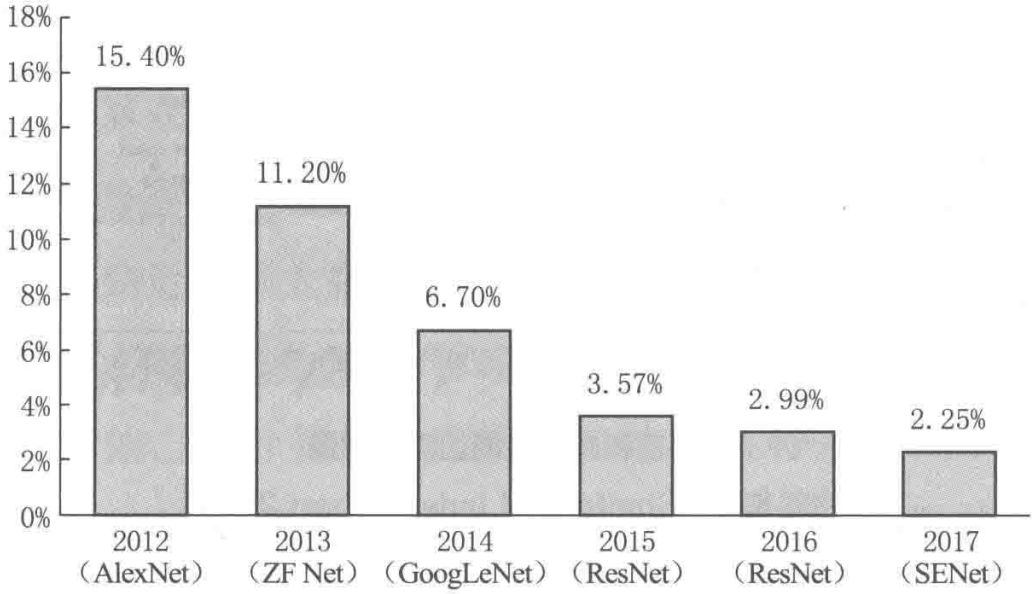
大语言和多模态模型的参数量

资料来源：Stanford AI Index Report 2023

② 深度学习的目前进展

深度学习算法意味着通过构建一种深层非线性网络

结构，来实现复杂函数逼近及自动特征提取，从而实现从有限的样本数据集中挖掘出一定的统计规律。作为机器学习的最新分支，深度学习的概念可以追溯到弗兰克·罗森布拉特（Frank Rosenblatt）于1958年发明的感知机（Perceptron），这种非常简单的单层装置是第一个人工神经网络。20世纪90年代以来，以杰弗里·辛顿（Geoffrey Hinton）、杨立昆（Yann LeCun）、约书亚·本吉奥（Yoshua Bengio）的深度学习模型为标志，深度学习技术突破了大型神经网络的研究瓶颈，取得了飞速发展，进入了技术成熟期，广泛应用于各种各样的实际场景。目前深度学习形成了四条发展脉络，分别为“计算机视觉和卷积网络、生成模型、序列模型、增强学习”。在技术成熟度方面，目前计算机视觉和语音交互尚处在技术应用初期，正开始尝试



2012—2017年ImageNet大规模图像识别竞赛最佳误差率

资料来源：image-net.org

不同的应用场景，其中计算机视觉是应用研究的重点方向。文本处理仍处于技术爬坡期，进展缓慢。

以计算机视觉领域为例，权威评测是 ImageNet 的大规模图像识别挑战赛（Large Scale Visual Recognition Challenge）。自 2009 年举办以来，大规模图像识别的错误率从 2009 年的 33.6%，已经下降到仅为 1.2%，错误率远低于人类的 5.1%，计算机视觉已经全面超越了人类。

③ 深度学习的未来趋势

2021 年 7 月，深度学习领域的三位图灵奖获得者约书亚·本吉奥（Yoshua Bengio）、杰弗里·辛顿（Geoffrey Hinton）、杨立昆（Yann LeCun）在《ACM 通信》杂志发表了一篇名为《人工智能深度学习》的论文，详细分析了当前深度学习模型所面临的挑战，指出未来深度学习模型可以在人类的帮助下进行学习，以提高应对环境变化的灵活性，解决广泛的反射和认知问题。

具体而言，当前人工智能的深度学习面临以下几方面的挑战：一是深度学习模型中的人工神经网络，缺乏生物网络的效率、灵活性和多功能性。二是无论是监督学习还是强化学习都存在缺陷，深度学习对他们的改进有限。监督学习离不开大量的人类标记数据，成本太高。强化学习只基于有奖励时才有用，而且需要大量的互动。三是深度学习系统能够解决的问题非常有限。只有在它们训练过的狭窄领域才有

效，哪怕轻微地超出边界，深度学习都会误入歧途。

对此，未来的深度学习将主要从以下几个方面突破：一是将神经网络与经典符号系统相结合，发展混合人工智能，解决深度学习在符号操作领域的短板。二是发展类似Transformer的神经网络架构。Transformer是OpenAI公司的ChatGPT-3和谷歌的Meena等语言模型的核心，可以在数据无标记的情况，通过无监督学习来开发表示。与卷积神经网络结合时，能够预测未知的内容。三是发展对比学习。即试图寻找缺失区域的向量表示，而不是预测精确的像素值。这种技术更接近人类的思维方式。四是发展胶囊网络。胶囊网络可以提供“直观物理”的深度学习，可以让系统理解三维环境。

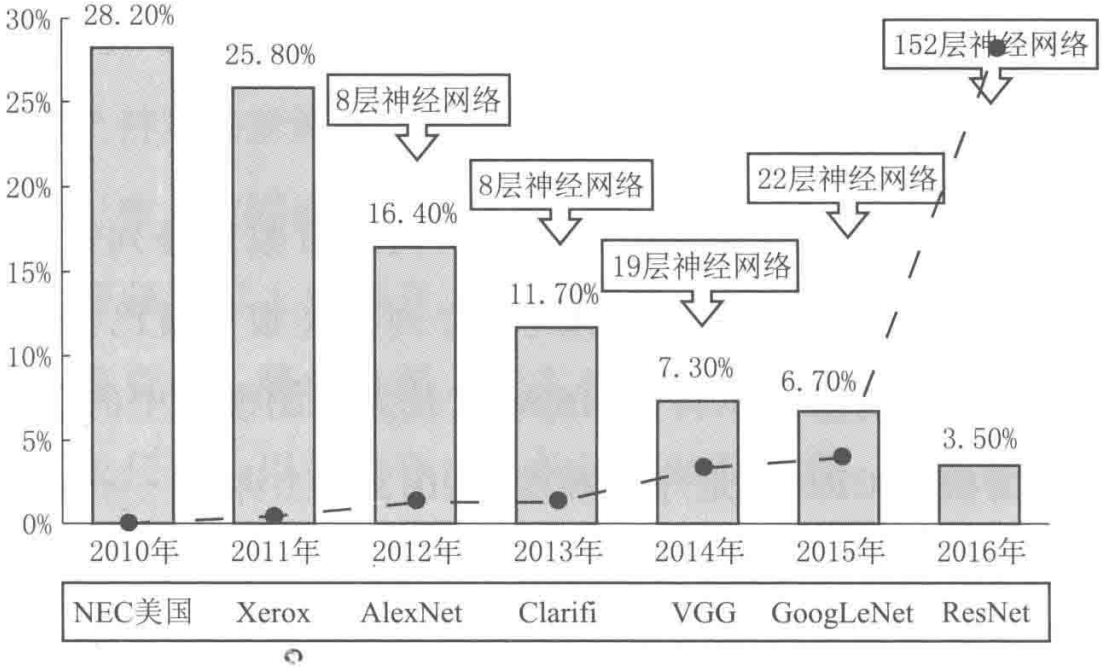
（2）深度学习的种类

① 卷积神经网络

计算机视觉和卷积神经网络是深度学习的第一个发展脉络，目前主流的应用多是使用基于多层网络神经的深度算法。卷积神经网络CNN最早可以追溯到休伯尔（David Hubel）和维泽尔（Torsten Weisel）于1962年对猫大脑视觉系统的研究。1979年，福岛邦彦（Kunihiko Fukushima）提出了卷积神经网络架构（Neocognitron），给出了卷积和池化的思想，开启了卷积神经网络的研究。在此基础上，杨立昆将BP算法应用到神经网络结构的训练，形成了当代卷积

神经网络的雏形。2012 年杰弗里·辛顿基于卷积神经网络 CNN 开发的 AlexNet 模型，在 ImageNet 竞赛中得了两个第一，正确率超出第二名近 10%，引起世界对深度学习的高度关注，并成为卷积神经网络的标配，被广泛发展出 VGG、GoogLeNet 等新模型。

现在主流卷积神经网络的应用主要基于多层网络神经的深度算法，一方面不断加强从海量数据库中自行归纳物体特征的能力，另一方面不断加强对新事物多层特征提取、描述和还原的能力，最终目标都是提高识别效率。研究表明 AI 错误率每下降一个数量级，应用范围就会上升一个数量级。错误率是决定行业被 GPT 革新程度的关键因素。以 ImageNet 的视觉竞赛为例，从 2010 年的 NEC 到 2016



2010—2016 年 ImageNet 图像识别错误率

资料来源：《人工智能行业研究报告》，36kr 研究院 2017 年

年的 ResNet，人类的算法模型由 NEC 的 33% 的错误率不断改进、迭代升级，到 ResNet 模型的错误率已经降低到 3.5%，最新的算法模型则降到了 1.5%，这意味着在视觉领域，基于卷积神经网络的机器视觉的图像识别能力已超越人类。

② 生成模型

深度学习的第二个发展脉络是生成模型。生成模型是机器学习领域的重要分支，但基于神经网络的生成模型却长期被忽视。2006 年，杰弗里·辛顿和学生一起设计了一个机器学习的生成模型，发表在《科学》杂志上，其利用单层的 RBM 自编码并且将其堆叠成为 Deep Belief Network，使用逐层贪婪或者 wake-sleep 的方法进行训练，使得深层的神经网络训练变得可能，这被认为是深度学习的一个开端。

③ 序列模型

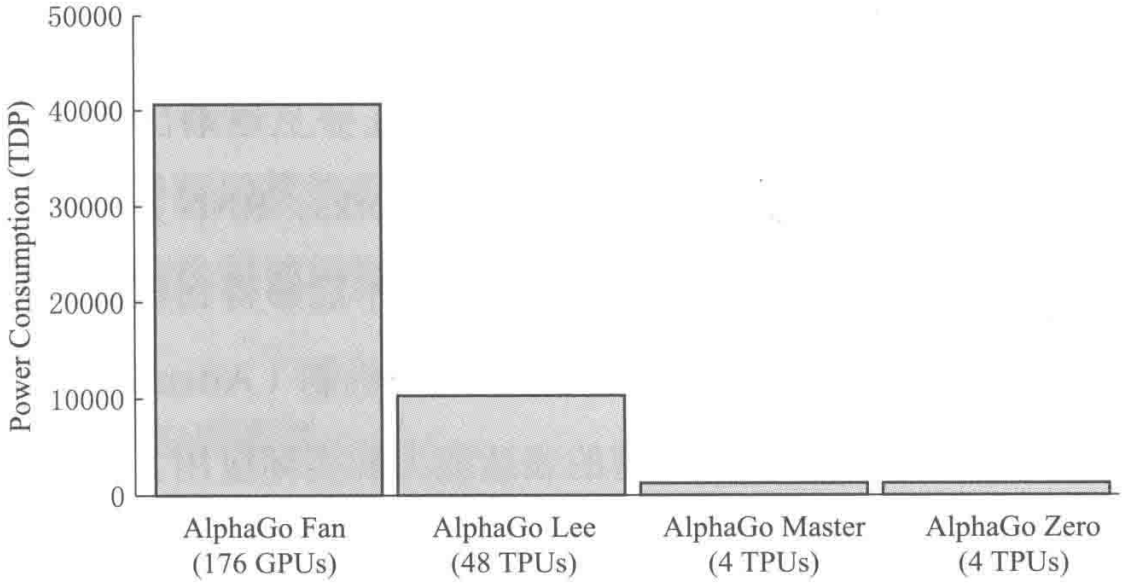
序列模型的原理是用过去的时间或者数据序列数据，去推导和判断未来的可能情况。序列模型很早就已提出，向量模型中的隐马尔科夫 HMM 以及无向图模型中的条件随机场模型 CRF 都是序列模型。2013 年杰弗里·辛顿使用序列模型研究语音识别的工作，其优于传统模型的表现，让序列神经网络模型引起广泛关注。约书亚·本吉奥随后提出了一种基于神经网络的语言模型，以及谷歌公司提出

了 word2vec（2013），利用反向传播的思想，较好地实现了文本翻译工作。后来，在机器翻译等任务上逐渐出现了以循环神经网络（Recurrent Neural Networks，RNN）为基础的 seq2seq 模型，通过语义信息压缩技术能够得到较高的文本翻译质量，这种方法也应用到注意力机制（Attention）的研究，大大扩展了序列模型的表达能力和实际应用。

④ 增强学习

深度学习的第四个发展脉络是增强学习（Q-learning）。这个领域最出名的企业是 DeepMind。2016 年 3 月，DeepMind 公司开发出阿尔法狗，该程序以 4:1 击败围棋世界冠军李世石，成为人工智能领域的里程碑事件。阿尔法狗应用的是增强学习中比较传统的 RL 算法，融入了传统的蒙特卡洛搜索技巧，2017 年又在此基础上提出了一个用阿尔法狗框架的棋类游戏的新算法，称为 Alpha Zero，取得了轰动世界的机器打败人类围棋冠军之成绩。

增强学习大幅提升了计算效率，但本身需要进一步节约计算资源、降低能耗，以实现较少的硬件资源完成算法的相关任务。以打败李世石的阿尔法狗为例，当时该系统使用了 1920 个 CPU 和 280 个 GPU，才能完成比赛任务，与人类冠军的一盘围棋比赛的能耗开支高达 3000 美元。谷歌此后在程序中引入 TPU 芯片，更新了自我学习的算法模



阿尔法狗不同版本所需的 GPU/TPU 资源

资料来源：雷锋网整理、平安证券研究所

型，使得模型的计算时间大幅缩短，只用 4 个 TPU 就完成了同样的计算任务，能耗显著下降、计算效率明显提升。

4. 各类通用技术

（1）训练和推断过程

人工智能系统的开发通常需要完成训练（Training）和推断（Inference）两个过程，以此开发出某些特定的能力。其中，训练过程是用已有数据进行学习，通过不断调整更新参数，使推断（或者预测）误差最小化的过程。推断过程是对新的数据进行分析，比如直接将数据输入神经网络并评估结果的正向计算过程，以此完成目标的分类、识别等特定任务。训练和推断在人工智能的技术研发中不可或缺，对语音识别、语义识别、计算机视觉等通用技术的发展非常关键。

训练和推断存在许多类似的基本运算，都需要使用大量的并行计算，对内存、带宽、网络延迟都有相似的要求。因此，在新基建中常常忽视两者的区别。但是由于两者在计算、存储方面存在显著的差异，尤其是两者的功能定位不同，因此现在越来越多的机构认识到搭建专用训练环境的重要性。也是说要搭建一个模拟现实的训练平台，专门训练人工智能模型，直至其成熟、稳定后，才能推向实际应用。而这个模拟训练的环境，其硬件资源的要求，甚至超过实际工作的平台。

对于人工智能的训练而言，需要更加重视计算的精度，因为训练的水平直接影响推断的准确度。支持训练的硬件必须支持具有较长字长的浮点数或定点数。另一方面，训练既有正向的计算，也有反向的计算，多重迭代计算对系统的算力提出很高的要求，不仅需要强大的计算能力，而且还要具备良好的扩展性，以应对数据量的暴涨。因此，训练的环境对于硬件的存储能力、计算能力、网络速度、平台的同步能力都提出了非常严格的要求，这甚至超过了实际工作中推理过程所需要的水平。这意味着我们在新基建的过程中，要同步甚至超前考虑人工智能的训练设施情况。

对推断过程而言，运算和存储的需求都远远低于训练。但推断应用场景更加复杂，从云到端的各种设备如数据中心，到各种智能终端如自动驾驶汽车、智能穿戴设备等，

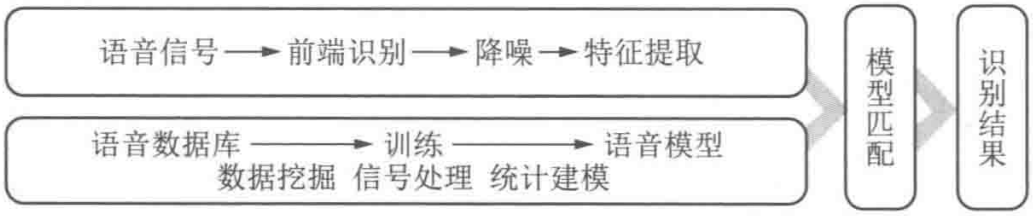
推断过程的需求和约束更具多样性。从需求出发，多数应用更关注速度、能效、安全和硬件成本，而模型的准确度和数据精度则视情而定。

虽然目前在理论研究中，机器学习方法清晰划分了训练和推断两个过程，但在增强学习（Reinforcement Learning）和在线学习（Online Learning）领域，模型一直处于持续学习和不断改进的进程中。因此，在未来的人工智能应用中，训练（学习）过程和推断过程会更多地交织在一起。

（2）语音识别技术

① 语音识别原理

语音识别是将语音转换为文本的技术，是自然语言处理的一个分支。前台主要步骤分为信号搜集、降噪和特征提取三步，后台会把提取的特征，交给语音模型对其进行解码，最终把语音转化为文本，实现让机器识别目标、理解语音。



语音识别技术架构

语音识别作为一种一维时域的信号，实现准确的语音识别主要有两个难点：一是数据的获取、清洗工作量极大。语音识别需要大量细分领域的标准化语料数据用于模型训练，收集各种地方语言，扩充语音库的声纹信息，是一件

任务量大且繁重的工作。二是语音特征的提取技术含量高。目前主要通过具备多层神经网络的深度学习来解决语义识别问题。多层神经网络的功能就是一个特征提取器，可对信号进行逐层深化的特征描述，最终从部分到整体，从笼统到具象，做到最大程度地还原信号原始特征。

科大讯飞的语音识别技术

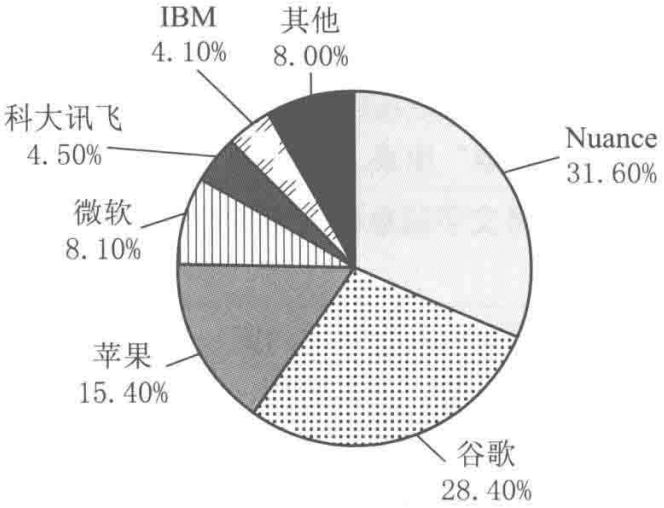
技　术		说　明
语音识别技术		让计算机能够“听懂”人类的语音，将语音中包含的文字信息“提取”出来。
语音合成技术		让机器将文字信息转化为声音信息，即让机器像人一样开口说话。
自然语言处理技术	基础研究	对自然语言内在规律的研究，主要包括词典编撰、分词断句、词性分析、语言模型、语法分析、语义分析、语用分析等。
	应用研究	面向不同的应用，研发相关的自然语言处理技术，主要包括拼音输入法、信息检索、信息抽取、自动摘要、机器翻译、语音合成、语音识别、文本匹配等。
语音评测技术		又称计算机辅助语言学习，机器自动对用户发音进行评分、检错并矫正。
声纹识别技术		通过语音信号提取相关特征，进而识别出说话人身份等，广泛应用于信息安全、电话银行、智能门禁以及娱乐增值等领域。
手写识别技术		让计算机能够“认识”用户在手写设备上书写的文字信息。

资料来源：科大讯飞官网

② 语音识别的市场

根据公开资料显示，目前语音识别的技术成熟度较高，

已达到95%的准确度，语音识别技术也已在智能手机中广泛应用，比如苹果iOS10.3版本中的Siri都已经能够识别上海方言了。当然，语音识别的准确度从95%到99%，才能带来质的飞跃，才能让人们从偶尔使用，到随时召唤，这也意味着语音识别的庞大市场目前还没有最后定型，各种科技初创公司在语音市场尚有拼杀空间。



全球智能语音企业市场份额

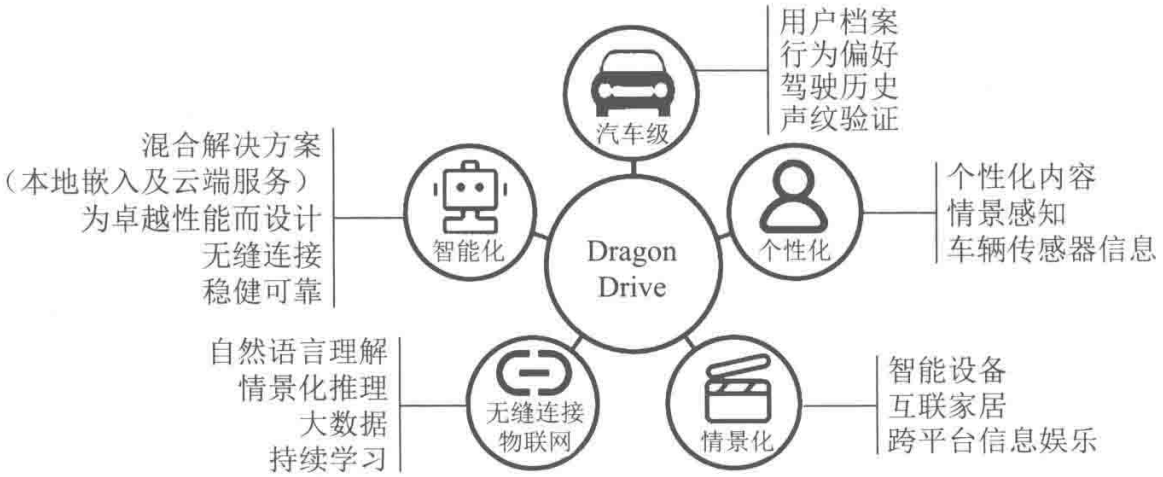
资料来源：Capvision，36氪研究院

据Research and Markets研究报告显示，近年来全球智能语音市场快速增长，2020年，全球语音市场规模达到191.7亿美元。根据Capvision报告显示，Nuance占据了全球语音行业市场份额的31.6%，谷歌占据28.4%，苹果占15.4%，国内霸主科大讯飞占据了全球4.5%的份额。

③ 语音识别的应用模式

随着全球范围内人工智能热度的高涨，语音识别作为其中重要的技术应用层落地项目也愈加多元化，国外和国

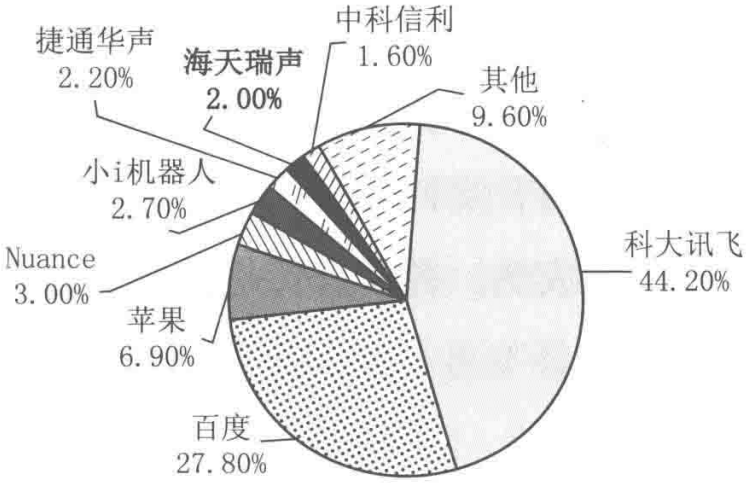
内也呈现出两种不同的技术应用模式：一是以 Nuance 的 Dragon Drive（声龙驾驶）为代表的互联汽车语音和内容平台模式。Nuance 是全球最大的独立语音识别公司之一，曾为苹果、三星提供语音支持服务，一度垄断了语音识别领域。随深度算法的成熟，各大公司纷纷开始自研语音识别技术，Nuance 的优势不再。但其互联汽车语音和内容平台“声龙驾驶”，仍为奔驰、宝马等众多知名车企采用。



声龙驾驶平台

二是科大讯飞的智能硬件研发模式。科大讯飞创办于 1999 年，主要从事智能语音及语言技术、人工智能技术研究，软件及芯片产品的开发，科大讯飞对下游的应用集成，主要是让开发商或客户自己完成。同时，科大讯飞公司目前在积极打造讯飞开放平台的人工智能生态，基于讯飞超脑、AI UI 内核，开发不同场景的全行业语音产品及服务，尤其是深耕智慧教育、智慧城市、智慧汽车、智慧医疗以

及智慧家居五个行业的规模化应用，取得了良好的收益。

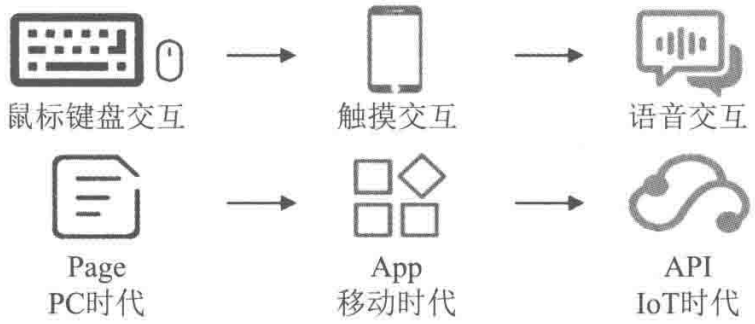


中国智能语音企业市场份额

资料来源：capvision

（3）语义识别技术

语义识别是自然语言处理（NLP）技术的重要组成部分，是人工智能的重要分支之一，主要解决“听得懂”的问题。



人机交互方式的演变

语义识别的作用主要是改变人机交互模式，将人机交互由最原始的鼠标、键盘交互转变为语音对话的方式。NLP在实际应用中最大的困难还是语义的复杂性，目前深度学习算法还不能算是语义识别领域的最优算法，也未能很好

NLP 通用架构示意图

资料来源：36 氪研究院

地解决语义识别问题。相信随着整个人工智能行业的加速发展，NLP 技术将会有长足的进步。

从 1996 年至今，自然语言处理无论在创业热度、获投数量还是获投金额都处于细分领域的前三。据 Global Market Insights 的数据，预计到 2024 年，语义识别市场规模达到 110 亿美元。当前，智能汽车是语音识别和语义识别的主要应用场景，有望成为最先爆发的领域之一。未来随着车联网向纵深化拓展，在硬件价格持续下降甚至趋于免费的趋势下，抓住语音交互这个天然流量入口，有望做强个性化增值服务，为企业创造更多的盈利点。这也是为什么各大企业在 Nuance 公司已经垄断了语音识别技术的情况下，还纷纷自主研发语义识别技术的主要原因。

语义识别领域的市场布局

公司	语义识别领域布局
微软	2008 年，收购语义搜索公司 Powerset，开始步入智能语音交互领域
	2014 年，微软推出全球首款语音助手 Cortana
	2016 年 6 月，收购语义分析公司 Wand Labs，以提升 Cortana 智能语音体验
	2017 年 1 月，收购了 Maluuba，强化了 Cortana 语义和文本理解能力
苹果	2010 年，收购 Siri 语音助手；2014 年 4 月，收购 Novauris；2015 年 10 月，收购 VocallQ
百度	2017 年 2 月，收购渡鸦科技，渡鸦团队并入度秘事业部

科技巨头如微软公司等，早在 2008 年就已开始布局语义技术领域。由于语义技术研发的复杂性，自主研发投入高、周期长，市场回报具有极大的不确定性，因此科技巨头们普遍选择了市场并购的模式，而非自主研发。这也为众多小型科技企业深耕细分场景鼓足了干劲。

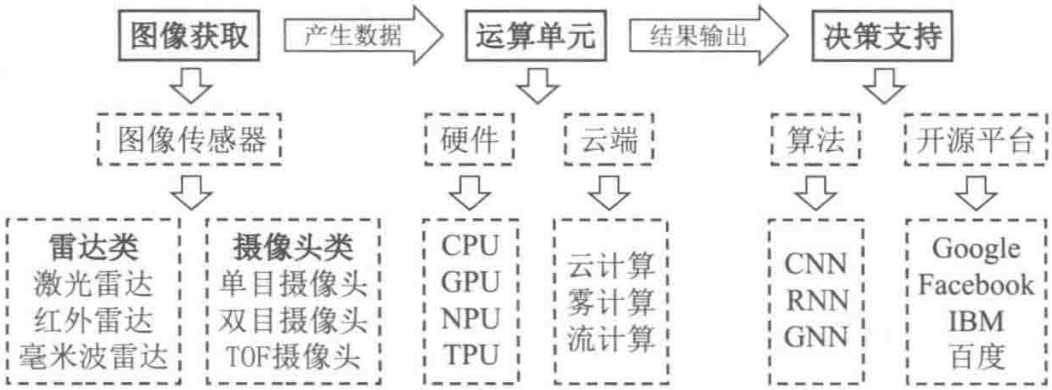
（4）计算机视觉

① 计算机视觉技术流程

前面在深度学习部分，我们曾提到卷积神经网络是计算机视觉（computer vision，简称 CV）的重要技术支撑，本节我们将继续深入探讨机器视觉问题。计算机视觉是指用计算机来模拟人的视觉系统，实现人的视觉功能，以适应、理解外界环境和控制自身运动。主要解决的是物体识别、物体形状和方位确认以及物体运动判断这三个问题。计算机视觉识别系统通常需要三个过程：目标检测、目标识别、行为识

别，分别解决了“去背景”“是什么”“干什么”的问题。

“去背景”是目标检测的过程，首先要通过传感器获取实时数据，少部分数据可直接使用传感器端的 MEMS 功能完成图像预处理，以减少数据传输的压力。大部分图像则需要对进行数据分割，继续传输至智能网络平台，由平台的智能“大脑”的运算单元和算法程序进行运算。“是什么”的过程称为目标识别。智能大脑通过“特征提取、目标分类、判断匹配”三个过程，完成对目标的识别。“行为识别”是则智能大脑对目标行为的分析与判断，通常可以分为模型建立和行为预判两个过程。



计算机视觉技术流程

② 计算机视觉的解决方案

上述去背景、是什么、干什么的过程，按照数据处理的环节不同，可以分为两类公司：一种是以商汤科技为代表的在后端用算法和学习模型解决问题的思路，着力推动视觉技术融入各行各业，打造人工智能商业生态。另一种是触景无限公司为代表的在前端用嵌入式智能硬件解决问题的思路。

商汤科技创办于2014年11月，是一家算法层的北京企业，专注于核心算法开发。通过自主开发原创深度学习模型，以视觉技术赋予计算机视觉感知和认知的能力。其自有的高性能算法库相对行业内开源平台库，较大程度提高了算法效率，带来2—5倍的性能提升。通过与京东、小米、新浪微博等应用层级公司合作，以及与MIT、斯坦福、香港大学、清华大学等高校及其实验室的科研人员，与谷歌、百度、微软、阿里巴巴等产业界的人员共同进行多媒体实验室合作，商汤科技大幅压缩了实验室研发成果与商业变现的时空差距，使得自己的算法可以很好地融合多类细分领域的特点，快速移植复制到各行各业。比如，商汤科技自身构建了具备200块GPU链接能力的DeepLink超算平台，过去耗时一个月的运算，现在只需5—6个小时即可完成。硬件门槛降低+计算能力提升，使得大部分企业快速接入计算机视觉技术成为现实。

触景无限公司代表了智能硬件的解决思路，其专注于前端嵌入式硬件并搭载软件智能硬件。基于英伟达、英特尔等嵌入式芯片，融合了人脸识别、物体识别、双目测距、GPS、TOF、IMU等多种传感器、数据处理平台和压缩算法模型，提供的嵌入式智能感知平台，具备体积小、功耗低、实时图像处理、处理能力高、多目摄像头支持、通用API接口等特点。例如，基于深度学习的人脸识别技术运

行在视觉卡上能达到每帧 100 ms 的速度（分辨率为 1080P 时的检测 + 识别速度），在 1:1 人证比对的情况下识别率大于 99%，在 1:50000 的情况下识别率大于 90%。通过该视觉卡，信息处理可以直接在前端完成，帮助前端硬件完成“感”与“知”的融合。

③ 计算机视觉行业的发展

计算机视觉应用场景可分为两大类：图像识别和人脸识别，每类又可继续划分为动、静共四个类别，基本覆盖了目前计算机视觉的各项应用场景。其中动态人脸识别技术是目前创业热度最高的细分领域，尤其是金融和安防场景，是其重点布局场景。

计算机视觉的应用场景

计算机视觉作为一种人工智能的基础技术应用，使用场景多样，市场潜力巨大。但由于各细分领域的成熟度相差较大，因此市场的应用和表现也呈现较大的差异。其中，人脸识别、指纹识别等生物特征识别领域，技术成熟度高，

工业化水平也相应较高。中国市场在其中占据了较大份额。

而在物体和场景识别方面，由于物体种类繁杂、形态多样，受图片质量、光照环境的影响，现有的图像识别技术难以解决图像残缺、光线过曝、图像过暗等问题。此外，受制于被标记数据的体量和数量，还很难实现突破特定应用场景的算法迭代问题，技术成熟度相对较低。

（四）应用层

1. 场景应用的流程

应用层主要支撑人工智能技术向各个领域场景应用，如医疗、金融、教育、农业、交通、安防等，其目的是用新技术改造现有产业，提高生产组织效率，提升服务质量，改善人类生活品质。

应用层技术更多偏重虚拟场景与外部各个行业既有知识和技术体系的对接、融合和改造。因此，应用层更多的是技术改造、产品升级、流程再造、组织变革的问题，技术问题与管理问题同在，甚至在人工智能技术应用的后期，更多是组织如何变革的问题。因此，在这里我们并不过多讨论细分行业的场景应用，只是就人工智能技术如何更好地应用到各个场景领域，结合技术应用到管理变革的四个过程，讨论如何做好人工智能的场景应用。

首先，人工智能技术作为一种前沿的技术突破，不是其他传统行业的人员所能详细了解的。人们通常只根据媒体、

网络以及自己相像，对人工智能如何在本部门、本单位应用，提出许多业务需求，其中既有合理的，也有许多脱离现实的，甚至还有许多推诿、甩锅的需求，这是在新技术、新模式推广应用过程中，非常普遍的现象。因此，推进人工智能场景应用的第一步不是技术问题，而是了解业务需求和场景环境。为人工智能技术的场景应用明确改造的目标，明确未来业务流程和技术环境。以此选择合适的技术和算法，从而能够更准确地进行定制化开发和定向优化。

其次，明确数据来源、处理方式和建模要求。媒体对场景应用的讨论更多属于典型案例的成果描述，而实际推进人工智能技术的场景落地，则需要从技术推进、产业变革和管理创新的角度，选择合适的数据来源、数据建模技术，包括数据清洗、特征提取、模型选择和数据训练等。

再次，选择合适的开发工具和人工智能模型。政府部门对人工智能技术还处于摸索、熟悉的过程中，面对市场上令人眼花缭乱的人工智能产品，难免无所适从。这里面需要考虑两个问题：一是未来的智能平台需要消除信息孤岛，人工智能应用需要尽可能实现统一平台、统一开发框架；二是信息安全问题。政府部门是各方关注的焦点，因此也必然成为各方技术攻击的焦点，人工智能目前还不完全成熟，国外在技术领先的前提下，已经在其中掺杂了许

多意识形态、信息误导、立场导向的内容，如何选择和部署人工智能的场景应用，需要处理好技术与管理的问题，不能重蹈技术公司控制政府信息部门的覆辙。

2. 场景应用的困境

人工智能已经展现了强大的功能和实力，人工智能技术的行业应用已是大势所趋。然而，我们也需要认识到人工智能技术并非万能，大规模的场景应用也会带来一系列的问题，突出有以下几点：

一是数据隐私和安全问题。人工智能需要大量的数据来进行学习和预测，掌握了数据也就掌握了用户的信息和隐私，企业的核心技术也不可避免地被他人控制。面对数据泄露和黑客攻击的风险，场景应用过程中，如何保护数据安全和用户隐私，如何保证企业不受制于软件公司，已经成为人工智能应用的一大难题。

二是人工智能模型的不可解释性。人工智能技术的发展，使得人工智能越来越像一个盲盒，人们输入数据后，并不清楚人工智能深度学习的算法是如何运作的，对其输出的结果无法事先预测，这不仅给人们带来恐慌，而且长期使用会削弱人类的思考能力、判断能力。当前，在短视频等媒体平台的信息推送环节，人们无法理解算法模型是如何做出预测和推送决定的，已经出现了严重的“信息茧房”问题。如果将这些技术进一步应用到生产制造领域，

人工智能的这种盲盒式的不可解释性和不可理解性，很可能给这些需要高度精确、明确解释的应用场景带来非常严重的后果。

三是道德和伦理问题。机器智能是否能够拥有情感问题，是人工智能领域的热点话题。人工智能技术向行业的深度应用，也必然引发一些道德和伦理问题，比如ChatGPT已经发现有说谎现象，当问及他不熟悉的问题时，ChatGPT不是诚实地回答不知道，而总是会给一些风马牛不相及的解释。虽然这一现象与程序设定和数据训练有关，但是也提醒人们，在人工智能的推广和应用过程中，需要建立起一套完善的伦理规范和约束机制。

四是人工智能是否需要为各行业定制不同的标准。场景应用的多样化、复杂化，必然带来标准和规范的多样化，这与统一智能平台的初衷并不相符。目前，业界提出了元宇宙的概念，用统一的虚拟场景空间，实现不同细分行业、细分场景的接入问题，但这并没有从根本上解决场景应用的可理解性、稳定性和可预测性等问题。尤其是不同的应用场景，需要使用不同的开发模块，必然存在成本高企的问题。是否需要从国家层面对此进行规范和统一，也是值得讨论的问题。

五是场景应用的法律风险问题。人工智能在细分领域的场景应用，实质是用机器智能替代人类智慧，这样就带

来了一个谁为行为后果负责的问题。目前，法律界对自动驾驶的讨论和研究较多，还有脑机接口、网络虚拟永生等问题，都是人工智能技术在场景应用领域，需要提前予以明确法律责任和义务的热点问题。

第四章

人工智能技术的应用与支撑

一、新基建、新平台

新一轮科技革命的第二个层次，是类脑技术创新所需要的软硬件支撑体系。基于这些物理架构的设施和设备，为新科技革命提供了实体支撑，我们将之统称为新基建、新平台。这里我们重点介绍政府部门比较关注的新基建。

（一）新基建的定义与属性

1. 新基建的定义

2018 年 12 月，中央经济工作会议确定 2019 年重点工作任务时提出“加强人工智能、工业互联网、物联网等新型基础设施建设”，这是新基建首次出现在中央层面的会议中。此后，中央政治局会议、国务院常务会议等也多次提到了新基建的具体内容。

2020 年 4 月 20 日，新型基础设施的范围正式明确，国家发改委在新闻发布会上指出：“新型基础设施是以新发展

理念为引领，以技术创新为驱动，以信息网络为基础，面向高质量发展需要，提供数字转型、智能升级、融合创新等服务的基础设施体系。”新基建主要包括三方面内容：一是信息基础设施，包括以5G、物联网、工业互联网、卫星互联网为代表的通信网络基础设施，以人工智能、云计算、区块链等为代表的新技术基础设施，以数据中心、智能计算中心为代表的算力基础设施等。二是融合基础设施，主要指深度应用互联网、大数据、人工智能等技术，支撑传统基础设施转型升级，进而形成的融合基础设施，如智能交通基础设施、智慧能源基础设施等。三是创新基础设施。主要是指支撑科学研究、技术开发、产品研制的具有公益属性的基础设施，如重大科技基础设施、科教基础设施、产业技术创新基础设施等。

2. 新基建的特征

与传统的基建设施建设相比，新基建具有以下几个特征：

一是突出面向数字经济。国家明确新型基础设施是“以信息网络为基础、技术创新为驱动”，重点面向数字经济领域建设。同时，更加强调核心网络与移动通信网络的协调发展，以此打造经济发展的新引擎。

二是突出提升拓展。新型基础设施不是全盘推倒原有基础设施，而是强调新型基础设施的“提升拓展”功能。

即提升互联网和移动互联网的服务能力；拓展则是在之前网络基础设施建设的基础上，拓展部署空间信息基础设施和物联网，从而培育这些领域的新业态、新模式、新产业。

3. 人工智能新基建的种类

相比传统的基础设施，新型基础设施建设突出科技创新驱动、数字化、信息网络三个基本要素。人工智能作为新一轮科技的主攻方向，人工智能本身就被定义为一种新型基础设施。因此，人工智能的基础设施自带新型基础设施的要素属性，在目前的阶段，则主要集中在信息基础设施方面。

人工智能的基础设施可以区分为算据、算力、算法三个层面。在算据层面，数据中心是人工智能的第一类新型基础设施。数据是新科技革命和产业变革中社会经济运行最重要的投入要素，在万物互联的智能经济环境下，来自IoT场景的实时数据需要重点关注，因为实时数据更具市场价值。这类数据对延迟敏感、密集度高，大部分需要在智能终端进行前缘数据处理，带动了边缘数据中心的崛起，未来将形成云计算与边缘计算协同发展的态势。

在算力层面，构建专用设施是人工智能的基础设施建设的核心。算力是涵盖数据收集、存储、计算、分析和传输的综合能力。在数字经济时代，国家的核心竞争力由算力来引领，包括计算速度、计算方法、通信能力、存储能

力、数据总量等。目前异构计算及其应用场景，特别是大算力应用场景逐渐清晰，成为建设的重点。一方面，在异构计算范式的场景环境中，人工智能开发涉及的软硬件环境愈加复杂，边缘计算的算力获得需要大量技术和基础设施投入。另一方面，深度学习和机器学习对大规模计算的算力要求则更胜一筹，巨型数据中心的建设成本和硬件资源耗费，是一笔极为高昂的开支。

在算法层面，根据人工智能产业的逻辑框架，通常我们把与算法相关的产业归入“塔尖的人工智能”一类。理论上，人工智能开发平台和软件框架可以借助算法模型，降低开发运维成本，其中算力和算法则共同决定数据要素转化为经济社会价值的效率和效果。目前中国人工智能产业主要依赖以 TensorFlow、Caffe 等为主的美国企业或机构研发的算法框架，而新基建强调加强自主创新，推动中国企业构建自主可控的算法支撑体系需要摆上议事日程。

（二）新基建的重点内容

1. 芯片工厂

芯片是整个人工智能革命的动力支撑，芯片工厂也因此成为人工智能新基建的首要项目。目前全球共有 104 家芯片工厂，主要分布在亚洲的中国、韩国、日本、新加坡，北美洲的美国，欧洲的英国、法国、德国。

2020 年全球前十大半导体设备供应商

排名	公司名	半导体设备营收 / 亿美元	全球市场份额
1	应用材料（Applied Materials INC）	163.65	17.7%
2	阿斯麦（ASM Holding NV）	153.96	16.7%
3	泛林半导体（Lam Research Corp）	119.29	12.9%
4	东京电子（TOKYO Electron Ltd.）	113.21	12.3%
5	科磊半导体（KLA-Tencor Corp）	54.43	5.9%
6	爱德万（Advantest）	25.31	2.7%
7	迪恩士（Dainippon Screen）	23.31	2.5%
8	泰瑞达（Teradyne）	22.59	2.4%
9	日立高新（Hitachi High-Tech）	17.17	1.9%
10	ASM 先进半导体材料（ASM International）	15.16	1.6%

全球半导体设备企业主要产品分布（2020 年）

排名	公　司	半导体设备销售 / 亿美元	CVD/PVD	光刻	刻蚀	离子注入	热处理	清洗设备	过程工艺控制	CP&FT 检测
1	应用材料	163.6	√		√	√	√	√	√	
2	阿斯麦	153.9		√						
3	泛林半导体	119.3	√		√			√		
4	东京电子	113.2	√		√			√	√	
5	科磊半导体	54.4			√				√	
6	爱德万	25.3								√
7	迪恩士	23.3			√			√	√	
8	泰瑞达	22.6								√
9	日立高新	17.2			√			√		
10	ASM 先进半导体	15.2	√							

其中，中国台湾有 17 座芯片工厂，占据全球芯片代工份额的五成以上。中芯国际有 17 座芯片工厂，产量占全球第五，但是一半以上是 45 nm 以上的，28 nm 以下量产的不多，14 nm 量产则落后台积电两代，基于 DUV 浸润式曝光的 7 nm 生产线 2023 年刚试产。英特尔有 15 座芯片工厂，规模全球最大，均为 45 nm 以下的先进制程。

2. 数据中心

（1）数据中心的分类

数据中心是各个行业运行信息系统的物理载体，是数字经济发展不可或缺的底座支撑，也是智能经济和社会运行的关键基础设施。近年来，伴随大模型对算力需求的快速增长，数据存储、运算和场景应用也呈现指数级增长，数据中心迎来了大发展的战略机遇期，年均增速高达 30%。

作为新型算力基础设施，数据中心有多种分类方式。通常根据功能定位，数据中心可分为三个层次：第一层次为通用型数据中心，主要容纳电子数据和电信系统。第二层次是 IDC（Internet Data Center），即提供互联网服务的数据中心。第三层次是提供云服务的数据托管中心，包括云储存、Laas、PaaS 等。根据服务对象不同，数据中心可以分为国家数据中心（NDC）、企业数据中心（NDC）和互联网数据中心（IDC）。根据规模的不同，数据中心又可以分

为超大型、大型、中型、中小型四类，其中超过 1 万个机柜的为超大型数据中心，5000—10000 个机柜的为大型数据中心，3000—5000 个机柜的为中型数据中心，小于 3000 个机柜的为中小型数据中心。

（2）数据中心的发展趋势

传统数据中心正加速与 5G 网络、云平台融合发展，向新型数据中心转型升级，总体表现为两个发展趋势，一是向低碳数据中心发展；二是向智能数据中心升级。

① 低碳数据中心

中国已明确到 2060 年实现碳达峰的远景目标，积极稳妥推进“碳中和”，实现绿色、可持续发展，将是各行各业的长期任务。作为新型基础设施，数据中心的能耗惊人，据《科学》杂志 2020 年发表的文章《重新校准全球数据中心能耗估算》的数据，2018 年全球数据中心总耗电达到 205 亿千瓦时，占当年度全球总用电量的 1%，节能降耗的空间巨大。

对于中国而言，发展低碳数据中心也十分紧迫。首先，中国的数据中心能耗依然偏高。截止到 2019 年年底，全国超大型数据中心平均 PUE（电能利用效率）为 1.46，而美国 Facebook 公司机房的 PUE 为 1.11。全球领先的数据中心运营商 Equinix 公司的 PUE 为 1.3。

其次，中国有能力提高数据中心的能效水平。比较而

言，中国数据中心能效水平近年来在不断提高，部分优秀绿色数据中心，如字节跳动公司的官厅湖大数据产业基地，大规模应用了间接蒸发等先进冷却技术，年均运行 PUE 达到 1.17，实现了全球领先。

再次，数据中心有望提升可再生能源的应用比例。展望未来，在数据中心的数量和规模都将持续扩张的情况下，提升数据中心的整体能效水平，也能为企业节省大量的开支。2021 年中国可再生能源发电量 2.49 万亿千瓦时，占全部发电量的 29.7%，有力地增进了能源供应安全，减少对进口化石能源的依赖。在价格方面，市场化绿电交易的价格优势在部分省市已经凸显，光伏发电在多数地区已经具备了与新建燃煤发电竞争的能力。性价比凸显吸引了大量互联网企业使用可再生能源，2018 年百度山西阳泉数据中心有 23% 的电力供应来自可再生能源。阿里巴巴张北云计算庙滩数据中心可再生能源使用占比达 50% 以上，年均 PUE 为 1.19，非常接近 Facebook 的水平。

② 数据中心模块化、智能化

作为核心的新型基础设施，数据中心本身就是一个资本密集、技术密集、高度市场化的高科技产品集合，管理和维护这样的智能大脑中枢，同样需要应用各类前沿的智能技术。针对传统数据中心的建设、运营和维护面临三项挑战：建设周期长、能耗高、运维效率低。未来的数据中

心需要能够实现低碳节能、快速部署、低成本建设运营，数据中心的预制化、模块化、绿色化、智能化因而成为未来的发展方向。

首先，数据中心的物理架构会日益极简化。传统的数据中心是攒机模式，各个子系统是彼此分离的。优点是各单元独立，相对安全；缺点是后期运维困难，管理成本偏高。对此，未来的数据中心将广泛应用模块化技术，从弱电设备、环境设备，直到核心的 IT 模块，模块化的解决方案和模块化硬件将遍布整个数据中心。极简架构 + 模块化，使数据中心能够快速部署、弹性扩容，同时大量的通用模块成本低廉、运维简单，能够满足高效节能的要求。

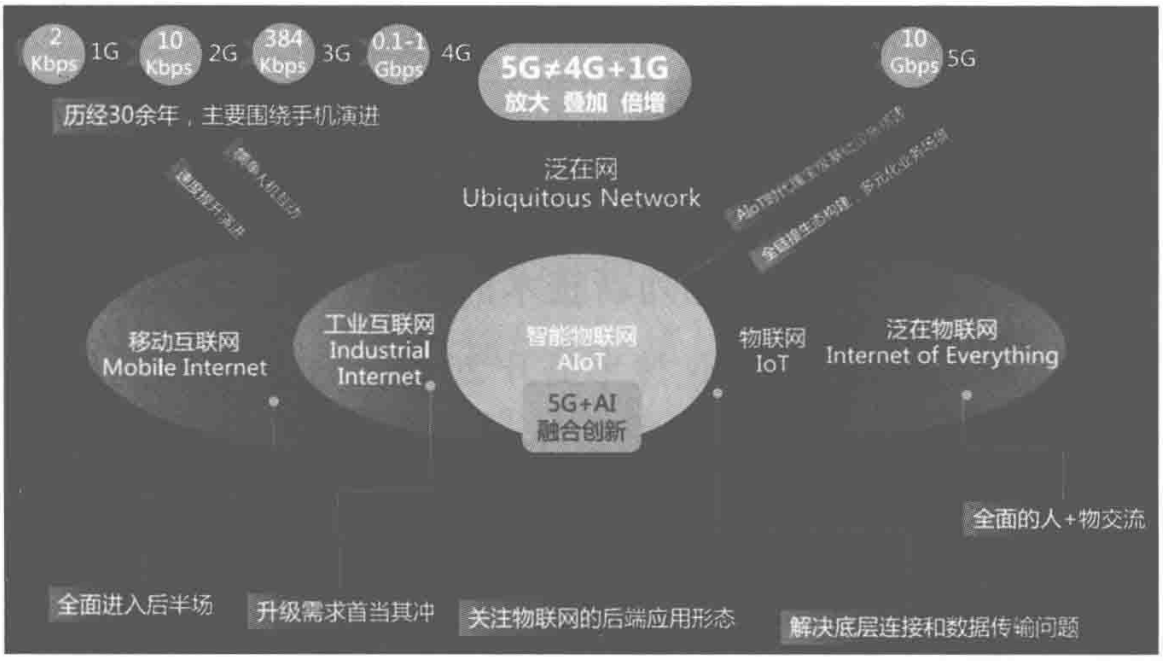
其次，数据中心的管理自动化、无人化。现有的数据中心虽然大量。依托最新的人工智能技术、远程通信技术，数据中心将进一步向无人工厂模式进化，未来除了核心的数据服务器无须人工干预外，整个数据的保障系统和附属系统，诸如供电系统、温控系统、自我检测系统，都将由智能系统、智能机器人自动完成，一个数据中心就是人类无人工厂的高端样板，集成着人类在智能制造领域的尖端技术。

3. 5G 和 6G 网络

（1）5G 网络

新一轮科技革命也是基于互联网的产业革命。用户、

产业、社会（CBS）三位一体，是这个时代科技进步与需求变迁的鲜明特色。正如智能手机的普及和 4G 网络的发展，带来了微信、短视频等新产品，改变了网络的沟通模式，人工智能搭配 5G+ 时代的卫星互联网，也必将跨界产生更加丰富的应用场景和崭新业态，比如空中高速上网、全息通信、智能制造、人体智能孪生、智能移动载人平台等。



智能物联网——5G 与人工智能的融合创新

① 5G 是巨大的技术进步

5G 是最新一代蜂窝移动通信技术，具有高速率、大带宽、低延时、广连接的技术优势。5G 并非独立的、全新的无线接入技术，而是以 4G LTE 为基础，充分利用和创新现有的先进技术，对现有无线接入技术的技术演进以及一些新增的补充性无线接入技术集成后解决方案的总称。

5G 移动通信的三大关键技术

技　术	内　容
基于 OFDM 优化的波形和多址接入	通过子载波间隔扩展实现可扩展的 OFDM 参数配置
	通过 OFDM 加窗提高多路传输效率
灵活的框架设计	可扩展的时间间隔
	自包含集成子帧
先进的新型无线技术	大规模 MIMO
	毫米波
	频谱共享
	先进的信道编码设计

资料来源：中国信通院

5G 不仅是若干成系列新技术的组合，更加是 4G 通信技术质的飞跃。一是通信速率提升 20 倍。5G 的峰值通信速率可达到 20Gbps，而 4G 只有 1Gbps。二是延迟改善了 10 倍以上。5G 的延迟小于 1 ms。三是可连接数提升 10—100 倍，IoT 应用可以强化。四是移动数据量将提升 100 倍。5G 具有超大带宽，远超 4G 网络。五是连接密度提升 10 倍，每平方公里可以连接 100 万终端。六是 5G 为建立全球物联网提供了可能。以 5G NR 统一空中接口为基础，人机共存的网络可以不断扩展，连接万物，最终构建起全球物联网。

5G 早在 2017—2018 年就开始成为热门话题，其跨界性、变革性吸引了众多投资。以 5G 为代表的新型基础设施不仅是数字经济的关键支撑，而且在服务垂直行业方面也

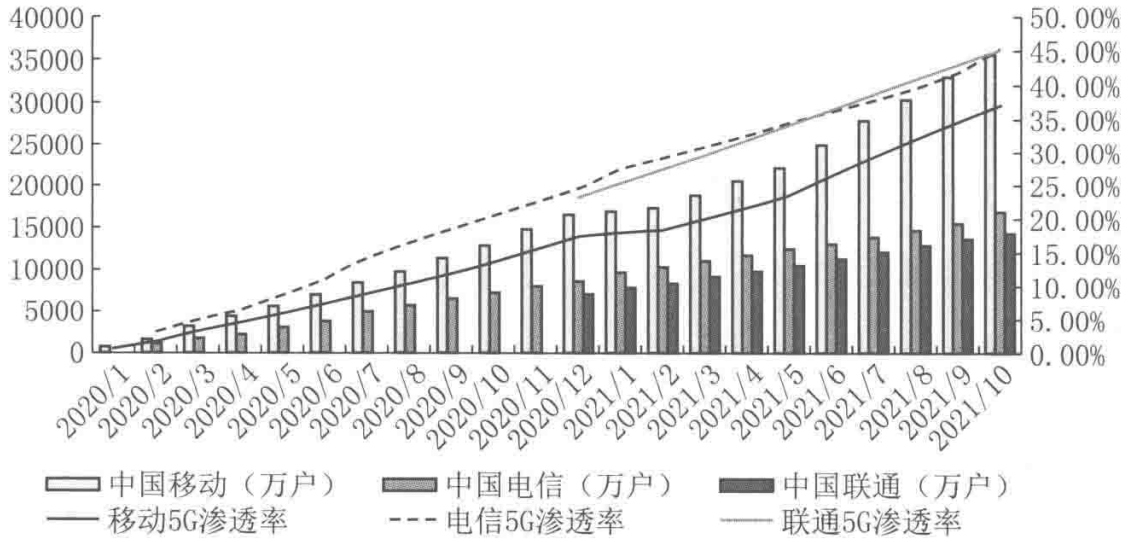
展示了相当的潜力。如果把 5G 网络看作一把钥匙，那么这把钥匙能够解锁场景数字化的难题，让数字技术以更小的颗粒度重塑现实世界。

② 5G 应用现状

经过数年的探索，5G 应用仍处于“样板示范”阶段。面对碎片化的行业需求，5G 仍未能较好地融合细分行业的业务需求，产生“爆款级”的场景应用，“应用场景缺乏”“投资巨大”“性能过剩”等负面评价，让 5G 的发展蒙上了阴影。比如目前产业界对超高带宽的需求并不强烈，仅有智能制造等高精尖的前沿领域需要 5G 网络。5G 的强大性能对于大量中低速场景，更是明显的性能过剩。在民用领域，4G 网络已能满足智能手机的场景应用，5G 网络的高昂运行费用和巨额的前期基础设施投入，并不能给用户带来明显的升级体验感。上述情况，都意味着 5G 的行业融合过程注定是一个渐进而漫长的过程。

自 2022 年下半年以来，随着国内 5G 规模化部署、5G 建设周期拉长、共建共享 5G 基站三重利好，运营商建设 5G 和运营维护的压力明显缓解，通信网络设备及器件成本不断下降，5G 渗透率不断提升。未来随着 5G 终端逐渐普及以及 5G 网络逐步实现深度覆盖，下游垂直行业应用将陆续实现质变，通信行业将从投资驱动 5G 建设，过渡到下游需求拉动 5G 繁荣的阶段。

事实上，近年来国内外对 5G 的讨论声音日益减少，而全球在 5G 领域的投资却在持续增长，这从另一个侧面让我们反思国外媒体否定 5G，与产业界真金白银投资 5G 的矛盾反差现象。仅就美国跳过 5G，直接大力发展 6G 的举动判断，5G 并非没有价值，而是因为中国在 5G 领域领先太多，而使得美国不得不改道走太空星链的路径。国内外的产业界都在埋头投资建设 5G 基站，恰恰说明了 5G 的重要性。因此，当前需要解决的是 5G 建成投产后的收支平衡问题和场景应用问题，以及 6G 如何与 5G 融合的问题，而非 5G 是否要建的问题。



国内三大运营商 5G 用户数及渗透率

资料来源：山西证券研究所

（2）卫星互联网（6G 网络）

① 卫星互联网的优势突出

卫星互联网是一种基于卫星通信技术的互联网，通过

发射特定数量卫星形成规模组网，向地面和空中终端提供宽带互联网接入等通信服务的新型网络。其中，地面站对用户发出的基带信号通过信源和信道编码、调制、变频后转换为射频信号（上行频率），放大功率后通过卫星天线发送到卫星处，再将卫星发射出来的射频信号（下行频率）进行低噪声放大、变频、解调等处理后形成基带信号，进而实现多个地球站间的通信。目前，最新的卫星互联网是马斯克的星链，已经通过发射特定数量的低轨道卫星形成规模组网，并在俄乌冲突中得到了实战检验。

卫星互联网的发展阶段

阶段	内容	主要代表	业务服务
第一阶段（20 世纪 80 年代—2000 年）	与地面通信网络正面交锋，展开竞争	铱星（Iridium）、全球星（Globalstar）、轨道通（Orbcomm）、泰利迪斯（Teledesic）、天桥系统（Skybridge）	窄带移动通信—低频段（低速语音、低速数据、物联网服务）
第二阶段（2000—2014 年）	作为地面通信的填隙和备份	新铱星（Iridium）、全球星（Globalstar）、轨道通（Orbcomm）	
第三阶段（2014 年至今）	与地面通信融合发展，扩展覆盖范围	Starlink、O3b、OneWebx、鸿雁、虹云	宽带互联网—高频段（高速率、低延时、海量数据）

资料来源：赛迪顾问、开源证券研究所

与传统高轨同步轨道卫星相比，低轨卫星互联网优势突出：一是组网后不存在通信盲区。与地面的 5G 通信网络

相比，低轨卫星通信系统拥有低损耗，低时延和高带宽的优势，多波束天线技术不仅能够提高卫星天线增益，还能覆盖更加广阔的地面区域，可以实现100%深度覆盖。二是6G网络的应用场景更加丰富。相比传统的高轨道通信卫星，新一代低轨道卫星互联网能够与地面组网实现空天地互联。可应用于高速/远距离移动终端，如飞机和轮船。应用场景包括偏远地区通信、海洋作业及科考宽带、航空宽带和灾难应急通信等。三是卫星互联网的建设成本更低。受益于各方技术的协同并进，低轨道卫星的部署成本大幅下降。比如，马斯克的SpaceX火箭回收技术基本成熟，提高了火箭的重复利用率，节省了巨额的发射成本，未来有望降到200万美元/次，而美国宇航局（NASA）此前的单次发射成本要超过40亿美元。四是星间链路的建设可降低对地面网络的依赖，减少地面段的复杂度和投资成本。

② 卫星互联网产业的构成

卫星互联网是一个全新的网络体系，其产业链由卫星制造、卫星发射、地面设备和卫星运营及服务构成。其中，卫星制造分为卫星平台和卫星载荷两大系统。平台系统保障基础工作，包括姿态及轨道系统、电源系统和推进、结构、遥测、热控等系统；载荷系统体现差异化功能，包括天线分系统和转发器系统。卫星发射包括火箭制造和发射服务；地面设备包括固定地面站、移动站和用户终端；卫星运营及服务

包含卫星移动通信服务、宽带广播服务和卫星固定服务。

卫星互联网的产业链构成

产业链	细分行业	行业组成
卫星制造	卫星平台	遥感测控系统、供电系统、结构系统、推进系统、数据管理系统、热控系统、资轨控制系统
	卫星载荷	天线分系统、转发器分系统、其他组件
卫星发射	火箭制造	推进系统、箭体制造、遥测系统、其他组件、发动机制造、制导和控制系统、安全自毁系统
	发射服务	火箭控制系统、逃逸系统、发射及遥测系统、发射场建设
地面设备	固定地面站	天线系统、发射系统、接收系统、信道终端系统、控制分系统、电源系统、卫星运控中心、卫星测控站
	移动站	集成式天线、调制解调器、其他设备
	用户终端	零部件、终端设备
运营及服务	卫星移动通信服务	移动数据、移动语音
	宽带广播服务	卫星电视服务、卫星广播服务、卫星宽带服务
	卫星固定服务	转发器租赁、管理网络服务

资料来源：赛迪顾问、开源证券研究所

③ 全球已展开卫星互联网竞赛

目前，美国 SpaceX 公司的星链（Starlink）计划已经成为全球卫星互联网的领头羊，预计共将发射 1.2 万颗卫星，实现全球 100% 的全覆盖。美国也正在积极推动星链与地面通信网络组成互联体系，相关工作主要由 OneWeb、SpaceX

等公司主导。随着星链的低轨道卫星互联网与地面通信网络融合一体，美国的6G网络有望达成更进一步的技术指标。其中，无论是空口延迟低于0.1 ms，还是毫米级的感知定位，TB级传输带宽，这些指标都将领先世界。

面对美国在卫星互联网的优势，2019年工信部成立了6G研究组，推动6G相关工作。2020年4月，国家发改委首次明确将卫星互联网纳入“新基建”范围，包括信息基础设施、融合基础设施和创新基础设施三个方面，因此2020年也被称为中国卫星互联网的元年。此后，各部委相继发布了多项产业促进政策，国内卫星互联网行业得到了快速发展。

中国鼓励卫星互联网发展的产业促进政策

发布时间	发布部门	政策名称	相关内容
2014.11.16	国务院	《关于创新重点领域投融资机制鼓励社会投资的指导意见》	鼓励民间资本研制、发射和运营商业遥感卫星，提供市场化、专业化服务
2015.10.26	国家发改委财政部国防科工局	《关于印发国家民用空间基础设施中长期发展规划（2015—2025年）的通知》	分阶段逐步建成技术先进、自主可控、布局合理、全球覆盖，由卫星遥感、卫星通信广播、卫星导航定位三大系统构成的国家民用空间基础设施
2016.12.27	国务院新闻办公室	《2016中国的航天》	建设天地一体化信息网络，基本建成空间基础设施体系，促进卫星及应用产业发展。培育卫星应用产业集群和应用市场，完善卫星应用产业链

续表

发布时间	发布部门	政策名称	相关内容
2018.4.19	国家发展改革委财政部	《关于降低部分无线电频率占用费标准等有关问题的通知》	调整网络化运营的高通量卫星系统频率占用费收费方式，降低开展空间科学研究的卫星系统频率占用费标准
2019.6.21	工业和信息化部	《卫星网络国际申报简易程序规定（试行）》	在卫星网络国际申报工作中建立简易程序，即境内卫星操作单位按照先申报后协调的原则，通过工信部向国际电联报送卫星网络资料，并开展国内协调等相应工作
2020.5.30	国家发改委	《关于2019年国民经济和社会发展计划执行情况与2020年国民经济和社会发展计划草案的报告》	支持商业航天发展，延伸航天产业链条，扩展通信、导航、遥感等卫星应用

目前，中国对卫星互联网的投资主要集中于空间段和地面段的基础设施建设，上游卫星制造、卫星发射及地面设备中的地面站建设都是优先发展的环节。随着组网成熟和基础设施环节逐渐落地，中游地面设备与运营环节将明显得到关注，下游应用服务也有望随之推进。根据赛迪智库测算，预计到2030年，中国低轨卫星互联网卫星制造环节将迎来投资高峰，中国卫星互联网整体市场规模可达千亿。

（三）新平台

1. 人工智能基础平台

智能经济就是基于人工智能的平台经济，新平台将成

为智能经济的基层组织。当前人工智能系统正由各自为政、各司其职的专用人工智能系统（弱人工智能），向高度集中的通用人工智能系统（强人工智能）演化，新平台发挥着促进各方良好沟通、协调运行和互相操作的作用，而且平台的专业知识和信息的流动效率反映了生态系统的整体水准，影响着整个生态系统的效率提升。

人工智能日前的快速发展，很大程度得益于各类平台的规模和效率，其中开源性平台的作用尤为显著。人工智能基础平台是人工智能产品由实验室成果转化市场合格产品的关键性训练平台，是探索、研发、调试、检证和完善各种各类智能算法、技术、产品和服务等的工具性系统，其作为连接底层硬件和上层应用之间的关键纽带，对人工智能的研究和产业发展起着决定性作用。传统的智能产品研发是一个异常艰苦的过程，不仅需要投入巨大的人力、物力，而且需要使用复杂的算法工程，做大量的细节维护工作，研发周期漫长，开发和维护费用高昂。人工智能基础平台就是一个集聚智力资源、计算资源、数据资源的平台，能够为海量训练数据提供自动标注，为智能产品开发和训练提供数据、算法、算力的协同优化，让人工智能学习模型从“手工”时代步入“自动”时代。

目前，国家比较知名的人工智能基础平台有国家发改委主导的“人工智能基础资源公共服务平台”，中国电科自

主研发的人工智能基础平台“电科芯云”。

2. 人工智能开放创新平台

2017 年 11 月 15 日，科技部启动了新一代人工智能发展规划和重大科技项目，宣布了首批共 4 家国家新一代人工智能开放创新平台，分别为：一是依托百度公司建设自动驾驶国家新一代人工智能开放创新平台；二是依托阿里云公司建设城市大脑国家新一代人工智能开放创新平台；三是依托腾讯公司建设医疗影像国家新一代人工智能开放创新平台；四是依托科大讯飞公司建设智能语音国家新一代人工智能开放创新平台。2018 年 9 月 20 日，依托商汤集团建设的智能视觉，成为第五家国家新一代人工智能开放创新平台。2019 年，依图科技、明略科技、华为、平安、海康威视、京东、旷视科技、360 奇虎、好未来、小米 10 家入选最新一批国家人工智能开放创新平台名单，中国的人工智能开放创新平台达到 15 家。

新一代人工智能发展规划，明确了构建开放协同的人工智能科技创新体系、建设安全便捷的智能社会、构建安全高效的智能化基础设施体系、加强人工智能领域的军民融合、培育高端高效的智能经济、前瞻布局新一代人工智能重大科技项目六大任务。在发展方向上，新一代人工智能开放创新平台聚焦人工智能重点细分领域，选择了大数据智能、跨媒体智能、群体智能、混合增强智能和自主智

能系统五个主力方向作为项目重点。力图从基础理论、支撑体系、关键技术、创新应用四个层面进行系统性、前瞻性任务布局。

2019 年 8 月，科技部印发了《国家新一代人工智能开放创新平台建设工作指引》，提出新一代人工智能开放创新平台是重要的创新载体，要充分发挥行业领军企业、研究机构的引领示范作用，有效整合技术资源、产业链资源和金融资源，持续输出人工智能核心研发能力和服务能力。《指引》鼓励人工智能细分领域领军企业搭建开源、开放平台，面向公众开放人工智能技术研发资源，向社会输出人工智能技术服务能力。

3. 云计算平台

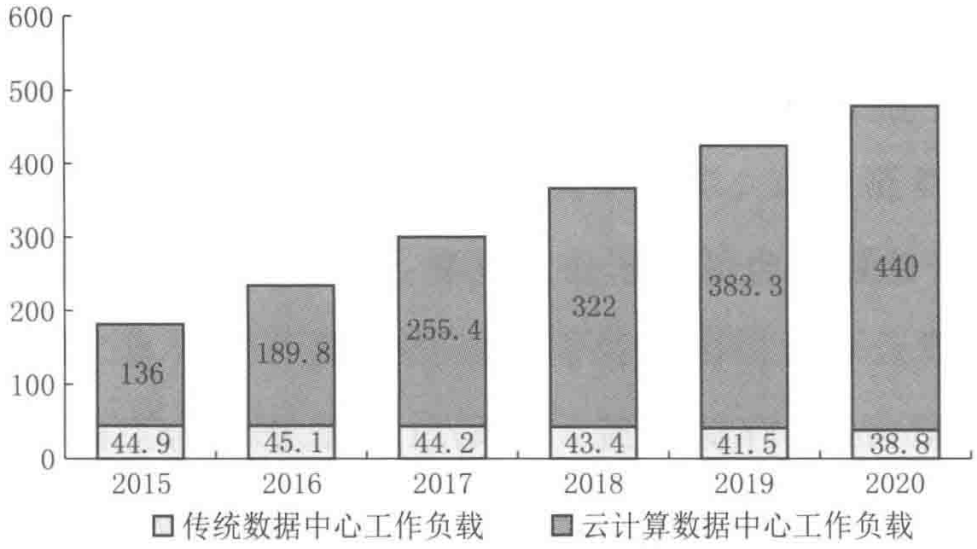
云计算平台也称为云平台，是人工智能算力的主要承载方式。云平台的核心是“云计算”的数据分享机制，能够为用户提供计算、网络和存储能力等软硬件资源服务。云平台的用户无须购买相应软件，只向“云”发出指示，就可以得到不同程度、不同类型的信息服务，而所有的计算任务都是在远程的云数据中心完成的。由于不需要自己建设数据中心、维护硬件、购买软件，因此企业和机构不再需要规划属于自己的数据中心，也不需要将精力耗费在与自己主营业务无关的 IT 管理上，使用云平台的成本大幅下降，灵活性显著提高。

云计算（Cloud Computing）是云平台的核心。云计算是一种新兴的商业计算模型，由分布式计算、并行处理、网格计算逐步发展而来的，目前仍处于发展之中，并没有一个权威的定义。与主要应用于密集型计算的超级计算机不同，云计算依靠其灵活的扩展能力，主要应用于社交网络、企业IT建设和信息化等数据密集型、I/O密集型的领域。

云计算平台可以划分为三类：一是以数据存储为主的存储型云平台，比如我们日常接触的百度云盘、EMC公司的Atoms云存储系统。二是以数据处理为主的计算型云平台，比如微软公司整合其软件及数据服务的Azure平台；亚马逊公司的弹性虚拟平台。三是计算和数据存储处理兼顾的综合云计算平台，比如谷歌公司的应用代管服务、Google AppEngine等。目前，中国云服务市场份额占比排名前四的是阿里云、华为云、腾讯云和百度云。而全球云平台市场占用率最高的是亚马逊AWS和微软Azure。

云计算之所以能够伴随着人工智能兴起，主要原因是其适应了开源平台的发展趋势，以及用户节省算力开支的需求。为此，开源的云平台将人工智能算法引入统一、公用的算力模型，并进一步将云计算与超级计算机技术结合，使计算服务能够做到密集运算与灵活扩展相结合，从而更好地适应不同行业的赋能需求。传统数据中心的工作

负载呈平缓的形态（Cisco 估计 CAGR 为 –3%），云计算数据中心的工作负载将占据绝大部分（Cisco 估计 CAGR 为 +26%）。同时云计算服务中心每台服务器的工作负载效率也明显高于传统数据中心（Cisco 估计为 3x）。



传统数据中心与云计算数据中心负载比较

资料来源：Cisco

云计算的出现有力地缓解了算力危机。根据南加大学者的研究表明，从 20 世纪 80 年代开始，全球服务器计算能力保持着 50% 的增长速度。但数据中心的整体开支并没有显著增加。其主要原因：一是摩尔定律的不断进步，数据中心性能提升的同时，成本逐年下降。二是标准的 x86 构架服务器普及化，带来了批量采购的价格优势。根据 IDC 估计，x86 服务器使用率已经达到了 82%。三是云计算使硬件资源虚拟化，带来的易隔离性、可扩展性，使资源得到充分利用，同时安全性也得到保证。

二、场景产业

（一）场景产业的特征

与其说场景产业，不如更准确地说是场景经济。如前所述，人工智能在应用层的主要功能就是场景应用。场景应用由元宇宙的虚拟空间统一接口、统一平台后，相应各个实体产业也需要顺应技术范式的转换，变身为新技术、新模式下的新型产业。因此，场景产业有两个来源，一个是人工智能技术向实体产业渗透所派生出来的场景产业，比如虚拟人、智能机器人、无人机等。另一个是实体产业转型升级后，将原有的产业场景化，由此衍生出来的场景产业。无论哪种模式，都会造就一种全新的场景经济。

一是行业跨界融合的趋势将会加强。人工智能、虚拟空间等新体系成型后，新场景、新产业会不断涌现，这就为不同行业的知识、元素交流带来了新的机遇，也因此会造就新的市场，比如疫情中我们使用的行程码，就是一个全新的场景应用。满足这些新的场景应用，就需要建设新的信息系统和业务体系，由此促进新产业部门的产生和原有产业链上下游的加速整合，相应政府部门的管理制度、政策规范也需要及时跟进调整。

二是新的经济模式将会持续创新。以无人驾驶领域为例，目前公众的关注焦点都集中在智能汽车、无人驾驶等

尖端领域，却没有注意到马路上的交警越来越少。事实上，由于人工智能技术的发展，中国已经在大中城市普及了天网系统和交通综合管理工程，视频监控、信息采集、电子围栏等一大批最新人工智能技术，被广泛应用到城市道路的管理之中，不仅派生出了大量的新兴产业，也极大地降低了城市道路交通管理的人力需求。类似的场景应用和场景产业，值得政府部门学习和效仿。

三是新兴产业和企业将构建新的行业生态。人工智能技术的普及催生了一大批新兴企业，这些企业聚焦于挖掘新的市场需求，提供新的智能产品和服务，打造新的消费模式，个性化定制、产品服务化、网络协同等新的产品模式、分工模式极大地颠覆了既有的产业竞争格局，为企业的创新、创业提供了新的行业生态，也为中国制造的再崛起、再出发提供了新生动力。

四是场景经济将成为城市竞争力的重要内容。当前，我们的城市综合竞争力仍主要关注道路、交通、基础设施、GDP 等指标，随着场景产业的崛起，未来打造营商环境将更加关注虚拟空间的场景产业和场景经济。当服务成为产品的主要内容和主要形态后，场景经济将不可避免地替代实体产品，成为经济交往的重要方面，这也提醒政府部门，需要在今后的工作，不仅要注重有形的硬件投资，而且要注重虚拟空间的交易和业务，做到软硬并举，同等重视。

（二）场景产业与生成式人工智能

随着 ChatGPT 的出色表现，生成式人工智能得到各国的高度重视，引发全球新一轮的人工智能投资热潮，场景产业也由过去“小作坊”式的碎片化开发模式，快速进入了以大语言模型为主导的“工业化”时代，并有望形成以若干底层大数据模型为支撑，众多“垂直数据 + 场景应用”为主导的产业生态。

1. GPT 加速推动应用场景开发

GPT 已经成为全球用户增长最快的应用，虽然目前在大多数应用场景中还暂时无法匹敌人类，但在一些特定的专业领域中，已经能够基本追平人类的表现。比如，GPT-3.5 在许多专业考试科目中的测试成绩已经能够超越绝大部分考生，在 GRE（国外研究生入学考试）中能够排名前 1%，在律师资格考试中成绩排名前 10%。

GPT-4 与 GPT-3.5 考试成绩对比

考　试	GPT-4	GPT-4（无视力）	GPT-3.5
统一律师考试（MBE+MEE+MPT）	298/400（约 90）	298/400（约 90）	213/400（第 10 位）
LSAT	163（第 88 位）	161（第 83 位）	第 149 次（第 40 次）
SAT 循证读写	710/800（约 93 路）	710/800（约 93 路）	670/800（第 87 位）
SAT 数学	700/800（第 89 位）	690/800（第 89 位）	590/800（第 70 次）

续表

考　试	GPT-4	GPT-4（无视力）	GPT-3.5
研究生入学考试（GRE）定量	163/170（第 80 次）	157/170（～ 62）	147/170（第 25 次）
研究生入学考试（GRE）口语	169/170（第 99 次）	165/170（第 96 次）	154/170（约 63）
研究生入学考试（GRE）写作	4/6（第 54 位）	4/6（第 54 位）	4/6（第 54 位）
USABO 半决赛 2020	87/150（第 99—100 次）	87/150（第 99—100 次）	43/150（31—33）
2022 年 USNCO 地方科考试	36/60	38/60	24/60
医学知识自我评估计划	75%	75%	53%
Codeforces 评级	392（低于第 5 名）	392（低于第 5 名）	260（低于第 5 名）
AP 艺术史	5（第 86—100 次）	5（第 86—100 次）	5（第 86—100 次）
AP 生物学	5（第 85—100 次）	5（第 85—100 次）	4（第 62—85 次）
微积分	4（第 43—59 次）	4（第 43—59 次）	1（第 0—7 次）
AP 化学	4（第 71—88 次）	4（第 71—88 次）	2（第 22—46 次）
AP 英语语言与写作	2（第 14—44 次）	2（第 14—44 次）	2（第 14—44 次）
AP 英语文学与写作	2（第 8—22 次）	2（第 8—22 次）	2（第 8—22 次）
AP 环境科学	5（第 91—100 次）	5（第 91—100 次）	5（第 91—100 次）
AP 宏观经济学	5（第 84—100 次）	5（第 84—100 次）	2（第 33—48 次）
微观经济学	5（第 82—100 次）	4（第 60—82 次）	4（第 60—82 次）
AP 物理 2	4（第 66—84 次）	4（第 66—84 次）	3（第 30—66 次）

续表

考　试	GPT-4	GPT-4（无视力）	GPT-3.5
AP 心理学	第 5 次（第 83—100 次）	第 5 次（第 83—100 次）	第 5 次（第 83—100 次）
AP 统计	5（第 85—100 次）	5（第 85—100 次）	3（第 40—63 次）
美联社美国政府	5（第 88—100 次）	5（第 88—100 次）	4（第 77—88 次）
美联社美国历史	5（第 89—100 次）	4（第 74—89 次）	4（第 74—89 次）
AP 世界历史	4（第 65—87 次）	4（第 65—87 次）	4（第 65—87 次）
AMC 10	30/150（第 6—12 次）	36/150（第 10—19 次）	36/150（第 10—19 次）
AMC 12	60/150（第 45—66 次）	48/150（第 19—40 次）	30/150（第 4—8 次）
品酒师入门（理论知识）	92%	92%	80%
注册侍酒师（理论知识）	86%	86%	58%
高级侍酒师（理论知识）	77%	77%	46%
Leetcode（简易）	31/41	31/41	12/41

资料来源：OPENAI 官网，《GPT4 技术报告》

同时，ChatGPT 虽然目前的优势仍主要集中于文字领域，但拥有强大的信息整合能力、自然语言处理能力，可谓是“上知天文，下知地理”，更具看齐人类的意识，已经能够稳妥地应对大范围、细粒度的逻辑问题，能够联系上下文给出创造性回答。最新的 GPT-4 引入了多模态处理能力，通过调整数据集、调整奖励函数等方式，能够进一步接收图片、视频等输入，甚至能够理解人类漫画中的幽默内容。

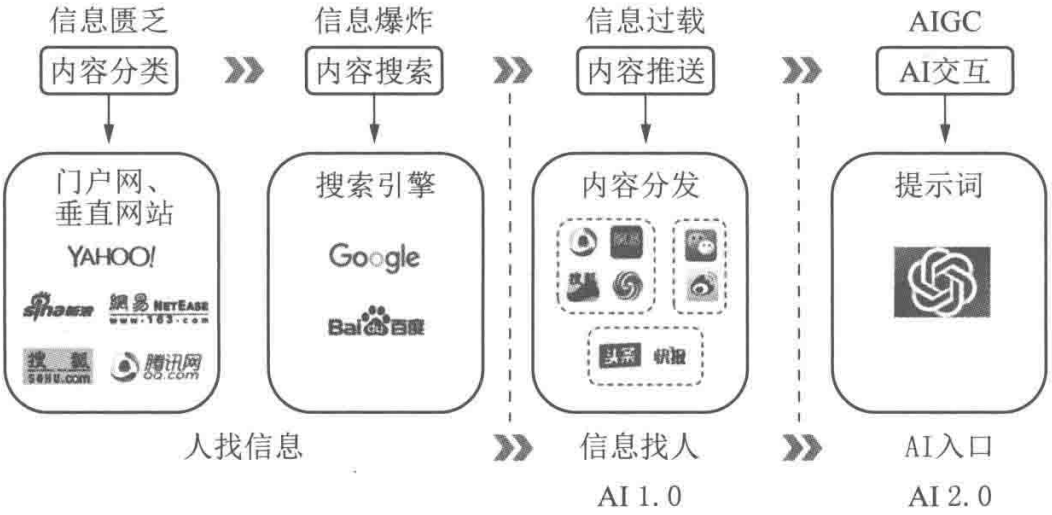
GPT 属于 Transformer 模型，作为特征提取器的效果要好于 CNN、LSTM 等模型，在自然语言生成的方面更出色。GPT 的强大性能还源于“自回归 +prompting”的算法架构，这种强大的功能也验证了“大模型预训练 + 小数据 fine-tuning”的有效性，只是对语义理解的精度不如 BERT 模型。与之相比，Transformer 模型虽然由谷歌公司最早提出，但谷歌公司选择了其中的 BERT 模型作为重点发展路径。BERT 模型同样也属于 Transformer 模型，但采用了双向 Mask 的方法进行训练，对自然语言的理解能力更好，比如百度公司推出的小度智能音箱，采用的就是 BERT 模型。

Transformer 模型重新统一了自然语言模型（NLP）的研究范式，机器智能得以使用单一的 NLP 模型应对不同的问题，无须像以前一样用不同的模型解决不同的问题，这大大缓解了模型参数堆积、资源分散的压力。而且 Transformer 模型在文字理解、图像识别等领域的表现优秀，性能超越人类水平，因此得以应用于更多的复杂场景。比如，特斯拉的自动驾驶项目采用了视觉识别系统而非雷达系统，其视觉系统的算法用的就是 Transformer 模型。与此同时，ChatGPT 还采用了强化学习和学习奖励的机制，改变了单纯堆砌参数、增加模型体积的开发思路，有效控制了算力开销和训练成本，为人工智能模型的发展和场景应用指明了方向，应用场景开发的“工业化时代”已经来临。

2. GPT 正在重构信息交互的场景模式

“工业化时代”的来临不仅源于通用算法模型的成熟，而且源于信息交互模式的升级。信息技术革命以来，人类的信息交互模式由“人找信息”进化到“信息找人”，目前正向“AI 交互”的场景模式演进，这将赋予场景产业全新的人机交互模式。

早在 20 世纪 90 年代，网络新媒体伴随着信息技术革命而崛起，催生了新浪、雅虎、腾讯等综合性的门户网站。这些门户网站以分门别类和外部链接的方式，为用户提供各类新闻和信息，其功能相当于把传统的报纸新闻、邮政服务搬到了网上，不同的网页标签就相当于报纸的不同版面，门户网站由此形成了对传统纸质媒介的替代，人们可以借助搜索引擎从网站上寻找自己关注的信息，是一种



人机交互方式的演变

资料来源：中信证券研究部

“人找信息”的网络交互模式。

2011 年以后，以微信、今日头条为代表的新一代移动互联网平台出现，重点解决了互联网用户普遍面临的信息过载和网络互信问题。其特色是通过用户认证、熟人圈子、内容推送等方式，在对终端用户进行精准画像的基础上，以内容的定向分发实现“信息找人”。但这种内容分发模式带来浓厚的“信息投喂”色彩，始终无法回避侵犯用户隐私的法律风险。同时，熟人圈子和信息投喂模式也带来的“信息茧房”等隐忧，越来越成为社会高度关注的热点话题。

而 GPT 的出现，创新了围绕提示词的人工智能交互模式，凭借人工智能模型的强大性能，弥补了此前信息单向流动的缺憾。人工智能交互为用户提供了一系列强大的功能，从需求分析、目标判断、信息搜索、关联信息、发散联想，这种相互交换信息的模式，不但更加符合人类沟通交流的习惯，而且能够为用户提供更加全面的深度信息和成体系的知识，甚至可以借助 GPT 来撰写论文、研发药物，这对人类的劳动用工模式、知识积累方式，乃至整个教育培训体系，都产生了颠覆式的影响。以人工智能交互作为元宇宙的入口，也必将开启一个全新的人机交互时代。

3. GPT 有望引导场景产业的未来

场景交互模式的变化，意味着 GPT 为人工智能产业

找到了商业化的着力点，新一轮的投资热潮也将持续提速。一是各国纷纷推出各自的大语言模型。自 2019 年 1 月 GPT-2 推出以来，其强大的性能得到了各界的广泛关注，各国陆续推出了相应的大语言模型和多模态模型。开放的模型架构、高频率的更新迭代，使得这些大语言模型的功能快速完善，能够满足多元化、碎片化的应用场景开发需求。二是场景应用的算力资源不再短缺。新基建项目的陆续投产，使得算力资源和算法模型的可获得性大大增加，规模场景开发的算力基础已经具备。同时模型的持续改进，也相应缓解了对算力需求的增长速度。三是各种大语言模型开始向 GPT 模型靠拢。

各国大语言和多模态模型情况

公布时间	国家 / 地区	研发机构	模　型	参数量
2019 年 2 月	美国	OPENAI 公司	GPT-2	17 亿
2019 年 5 月	美国	华盛顿大学	Grover-Mega	1.24 亿
2019 年 9 月	中国台湾	英伟达公司	Megatron-LM	
2019 年 11 月	美国	谷歌公司	T5-11B	11 亿
2020 年 1 月	美国	谷歌公司	Meena	26 亿
2020 年 2 月	美国	微软公司	Turing NLG	170 亿
2020 年 5 月	美国	OPENAI 公司	GPT-3	1750 亿
2020 年 8 月	中国	百度公司	ERNIE-GEN	10 亿
2021 年 1 月	美国	OPENAI 公司	DALL-E	12 亿
2021 年 2 月	中国	中国人民大学	WuDao-WenYuan	100 亿
2021 年 2 月	美国	OPENAI 公司	GPT-Neo	

续表

公布时间	国家 / 地区	研发机构	模　型	参数量
2021 年 5 月	中国	盘古鹏程公司	盘古 -alpha	20 亿
2021 年 5 月	中国	清华大学、阿里	CogView	40 亿
2021 年 6 月	韩国	Naver 公司	HyperClova	2040 亿
2021 年 6 月	中国	人大、北京智源	WuDao 2.0	1.75 万亿
2021 年 7 月	中国	百度公司	ERNIE 3.0	2600 亿
2021 年 7 月	美国	OPENAI 公司	Codex	120 亿
2021 年 8 月	以色列	AI21 Labs 公司	Jurassic-1-Jumbo	1780 亿
2021 年 10 月	美国	微软、英伟达	Megatron-Turing NLG 530B	5300 亿
2021 年 11 月	美国	DeepMind 公司	Gopher	2800 亿
2022 年 2 月	美国	DeepMind 公司	Alphacode	4000 亿
2022 年 3 月	美国	DeepMind 公司	Chinchilla	700 亿
2022 年 3 月	美国	谷歌公司	PalM(540B)	5400 亿
2022 年 3 月	美国	OPENAI 公司	Dall-E2	35 亿
2022 年 3 月	美国	Stability AI 公司	Stable Diffusion	66 亿
2022 年 4 月	美国	Meta AI 公司	OPT-175B	1750 亿
2022 年 4 月	以色列	AI21 Labs 公司	Jurassic-X	1780 亿
2022 年 4 月	美国	谷歌公司	Imagen	20 亿
2022 年 7 月	美国	谷歌公司	Minerva	5400 亿
2022 年 8 月	中国	清华大学	GLM-130B	1300 亿
2022 年 12 月	法国	科学家联盟	BLOOM	1760 亿

目前的算法大模型有两种发展路径：一种是 Open AI 公司 GPT 模型，采用了自回归 +Zero/Few Shot Prompt 路径。另一种是谷歌公司的 BERT 模型，采取了双向预训练 +Fine

Tune 的路径。GPT 模型牺牲精度换来了更理想的终端接口，因而无需区分任务的类别，更适合大规模部署。而谷歌公司的 BERT 模型，Fine-Tuning 的路径精度更好，但对于公司要求也更高，小公司更难部署。因此，鉴于 GPT 强大的产品化能力，未来的场景开发都将转向 GPT 的算法模型和技术路线。

（三）生成式人工智能的未来

1. 搜索引擎不会被完全替代

Chat-GPT 只是 GPT 的改进模型，回答问题的准确性和逻辑性有了较大提高。回顾 GPT 的发展历程，可以发现 GPT 作为生成式人工智能的代表，目前已经基本穷尽了英文世界的公开数据，GPT-3 的模型参数量达到令人惊叹的 1750 亿之多。随着人工智能向多元化的应用场景渗透，人工智能不得不借助搜索引擎的强大功能，去搜寻各个细分场景的加密信息。搜索引擎的机制是发散性的，与生成式人工智能的运行机理完全不同，在可以预见的未来，两者之间是融合和互补关系，而非替代关系。

需要引起注意的是，尽管 GPT 的性能取决于其受训练的数据输入，本身不可能输出超过自身“学会”的内容，但目前已经发现了 Chat-GPT 存在“胡言乱语”等问题，聊天机器人会非常自信地输出不真实的回答，这使得人类不敢放手使用 GPT。如何解决“AI 幻觉”和“虚构”等问

ChatGPT 发展历程

模型名称	GPT-1	GPT-2	GPT-3	InstructGPT	ChatGPT	GPT-4
发布时间	2018 年 6 月	2019 年 1 月	2020 年 5 月	2022 年 1 月	2022 年 11 月	2023 年 3 月
参数量	1.17 亿	15 亿	1750 亿	13 亿	未公布	未知
模型特色和思想	无监督学习。其设计思想是通过无标签的数据，先学习一个通用的语言模型，然后不断对其进行微调，进而处理有监督的任务	多任务学习。目标是训练一个泛化能力更强的向量模型。使用的是无监督的预训练模型。其验证了可以通过海量数据和大参数训练出来智能语言模型	海量参数。其设计的核心思想是 incontext learning（元学习）。Gpt3 借助网络的力量大幅提升了参数量，取得了优秀的表现	引入 PLHF，通过微调以遵循用户的意图，大幅提升了模型的泛化能力	广泛连接了自然界和真实社会的大量语料库，支持多轮对话，人机交互能力大幅提升，准确度更高	100 万亿参数，进一步解决了安全性、多模态问题，提供了图片识别功能

资料来源：作者根据公开资料整理

题，已经成为当下研究的热点。

2. 生成式人工智能将向多模态发展

在 GPT-4 已经大幅提高精度和准确性的情况下，未来继续增加模型参数量对模型准确性的收益递减。如前所述，从 GPT-3 开始到各国开发的各种大语言模型，网络中可获得的公开语言数据，如学术著作、文学作品、网站新闻、维基百科、社交媒体等资源，基本已经被充分利用，未来的智能产品将开发和利用各式各样的细分场景，这意味着生成式人工智能将向多模态的方向发展，场景化、定制化将成为主流。这也意味着，当前的人工智能将进一步升级通用问题的求解能力，向超级人工智能进化。伴随着算法模型复杂度的增加，数据源、工程能力与资本投入将决定这场超级人工智能竞赛的未来格局。同时，随着超人工智能的出现，机器人全面替代人类的问题也浮出水面，这是当前应用场景开发、人工智能升级需要提前考虑的重大议题。

第五章

人工智能革命的全球竞争

尽管全球人工智能技术远未成熟，但政府和产业界都在积极推动人工智能技术的场景应用。目前，随着人工智能技术的广泛应用，不仅在算法模型和人工智能芯片等核心技术领域，积累了强大的技术创新，而且进一步引致强劲的产业化、商业化动力。产业化带来了巨大的商业利益和深远的战略影响，促使大国纷纷加入人工智能的全球竞争。

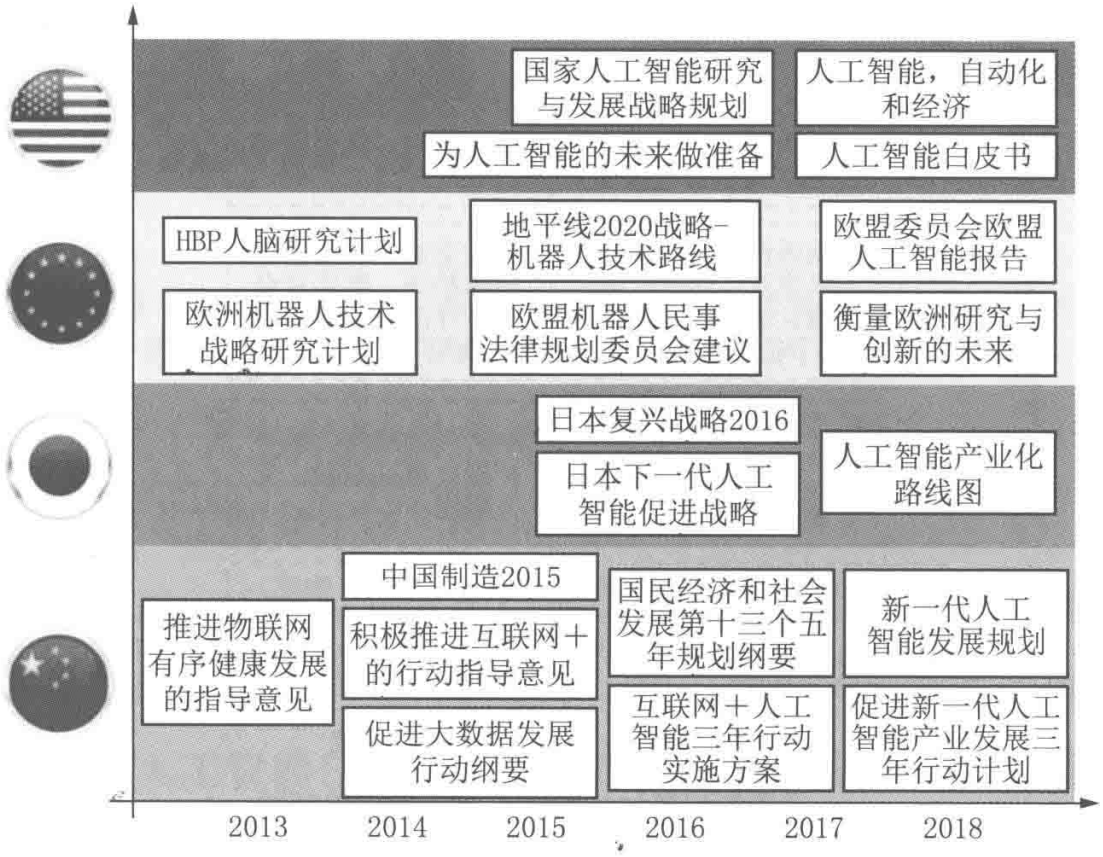
一、人工智能的全球竞争态势

近年来，人工智能对经济社会的巨大影响日益显现。尤其是2016年谷歌阿尔法狗战胜李世石之后，人工智能的强大潜力为全球所认识。在人工智能应用受到市场热捧的同时，越来越多的国家开始把握趋势，着手研究和制定国家层面的人工智能战略和发展规划，积极推动人工智能科

技的产业化，全球人工智能领导者之争已经开启。

（一）全球人工智能产业的政策布局

在人工智能的全球竞争中，主要发达国家始终走在技术创新的前沿，目前已经形成了中美两强竞争的格局，而日本、欧盟、韩国等则在各自的细分领域里占据着特殊的竞争优势。为了与中美两个大国竞争，加拿大、日本、新加坡、阿联酋、芬兰、丹麦、法国、英国、欧盟委员会、韩国和印度都发布了促进人工智能应用与开发的战略，这些战略关注点十分广泛，涵盖了科学研究、人才培养、技能与教育、道德包容、标准与法规及数据与数字基础设施



主要国家的人工智能发展计划

等诸多领域，而中国和美国面对这样的竞争态势，一方面加紧人工智能技术的研发，另一方面积极推动人工智能技术的场景应用。

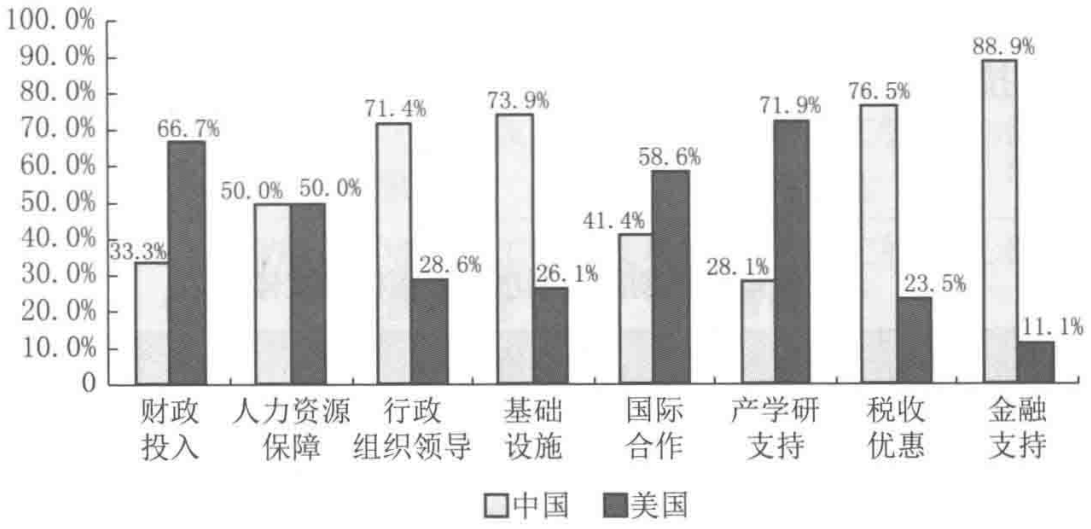
比较各国的政策，不难发现人工智能政策在技术领域的关注重点主要有四个方面：一是高度关注人工智能的基础理论和关键技术，尤其是技术的跨界应用。二是人工智能的感知技术。语音识别、大数据分析等是各国普遍关注的热点。三是超级计算与数据安全。一方面是对算力的追求，另一方面是对数据要素安全的高度重视。四是对通用人工智能的关注。神经元理论、感知理论、人机协作技术等，都是各国高度关注的领域。

技术领域	应用领域
数据分析，感知理论，广义人工智能，类人神经元，人机协作技术	国土安全，国防军事，医疗健康，智能交通，农业
数据保护，网络安全，人机交互，计算机识别，超级计算机，高性能计算，人工智能伦理	智能交通，医疗护理，金融经济，数字社会，生态经济，农业，能源
机器人，脑信息通信，语音识别，语言翻译，社会知识解析，大数据分析	制造，物联网，医疗护理，自动驾驶，无人配送
聚焦人工智能基础理论和关键技术，同时支持对人工智能交叉学科研究的探索	制造，农业，物流，金融，商务，家居，教育，医疗，城市管理，环境保护

人工智能政策关注的重点技术与应用领域

资料来源：华夏产业研究院（2018）

在应用领域，政策关注的门类非常广泛，基本涵盖了所有行业分类，体现了人工智能的普适性和易用性。从人工智能在我国的应用领域来看，人工智能的技术应用更多由资本主导，更加关注市场回报、市场占有率等短期指标，因此主要是各类弱人工智能工具的行业应用。而美国对于人工智能的发展，更加侧重强人工智能的通用性。因此，能够推出 ChatGPT 这样划时代的强人工智能产品。



资料来源：汤志伟、雷鸿竹、周维，《中美人工智能产业政策的比较研究——基于目标、工具与执行的内容分析》，《情报杂志》2019 年第 10 期。

（二）人工智能的中美比较

1. 人工智能比较的指标体系

当前人工智能对各个行业的渗透性、扩散性和颠覆性与日俱增，以预训练大模型等为代表的新兴技术正在引发新一轮人工智能创新热潮，各国都在积极跟踪人工智能的技术前沿，大力推动智能技术的场景应用，以期掌握智能

经济时代的主动权。在这股浪潮中，我国在人工智能领域已初步形成了应用导向、场景驱动、技术支撑的发展生态。但业内的普遍认为我国在底层算法、关键软硬件和生态体系等方面与国际先进水平还存在着较大的差距。

在全球人工智能的技术创新和产业竞赛中，中美两国暂居第一阵营。王山、陈昌兵（2023）根据技术创新周期的不同阶段特征等因素，从基础研究、技术创新、科技布局、产业规模与技术进步5个维度，对中美两国的人工智能技术创新发展水平进行了测度，提出了如下指标测度体系：[1]

人工智能技术创新发展水平多指标测度体系

维度	指　标	指标描述	含　义
基础研究	科技论文发表数	人工智能技术领域科技论文逐年发表数	反映基础研究的成果和动向
技术创新	优先权年专利申请量	年度优先权专利申请的数量	技术创新信息的重要结晶和载体
	优先权年IPC数	年度优先权专利国际专利分类号的数量	侧重功能分类，包含不同IPC小类的数量，表明技术覆盖范围
	优先权年MC数	年度优先权德温特手工编码（Derwent Manual Code）的数量	分类更加精准，包含不同MC小类的数量，表明技术覆盖范围

[1] 王山、陈昌兵：《中美人工智能技术创新的动态比较——基于人工智能技术创新大数据的多S曲线模型分析》，《北京工业大学学报》（社会科学版）2023年第3期。

续表

维度	指　标	指标描述	含　义
科技布局	研究与开发机构（R&D）数	人工智能技术所投入的研究与开发机构数量	反映人工智能技术研究与开发机构的布局状况
	研究与开发（R&D）人员	人工智能技术所投入的研究与开发人员数量	反映人工智能技术所研究与开发人员的布局状况
	研究与试验发展 R&D）项目数（课题）	人工智能技术所投入的研究与试验发展（R&D）项目数	反映人工智能技术研发项目布局状况
	研究与开发（R&D）经费投入	人工智能技术所投入的经费	反映人工智能技术研发经费投入布局状况
产业规模	企业数量	人工智能技术相关企业数量	考察人工智能技术市场应用情况
	新增人工智能企业数	新增的人工智能技术相关企业数量	考察人工智能技术市场吸引力
	上市企业数量	上市的人工智能企业数量	考察人工智能新技术市场关注度
	人工智能企业融资规模	人工智能企业融资规模	考察人工智能新技术融资吸引力
	市场规模	人工智能企业市场规模	反映人工智能新技术投资吸引力
技术进步	芯片	表征集成电路尺寸大小的一个参	考察集成电路小型化的能力
	GPU 算力	每秒所执行的浮点运算次数	表征人工智能芯片的计算能力

资料来源：王山、陈昌兵（2023）

借用这一指标体系，本书重点选取基础理论、技术创新、产业规模等三个维度中的部分关键指标，对中美在人

工智能领域的发展现状进行比较。

2. 基础研究的中美比较

在基础理论领域，可以从两个方面比较中美两国的人工智能创新进展和理论积淀：一是中美两国的人工智能技术创新指数。二是期刊和会议论文的比较。在技术创新指数方面，可以根据《人工智能技术创新发展水平多指标测度体系》的五维度指标，结合各国人工智能技术所处的发展阶段，对相关的指标赋予不同的权重，进而拟合各国人工智能技术进化的 S 型曲线，以此进行技术创新逐年发展指数的跨国比较。

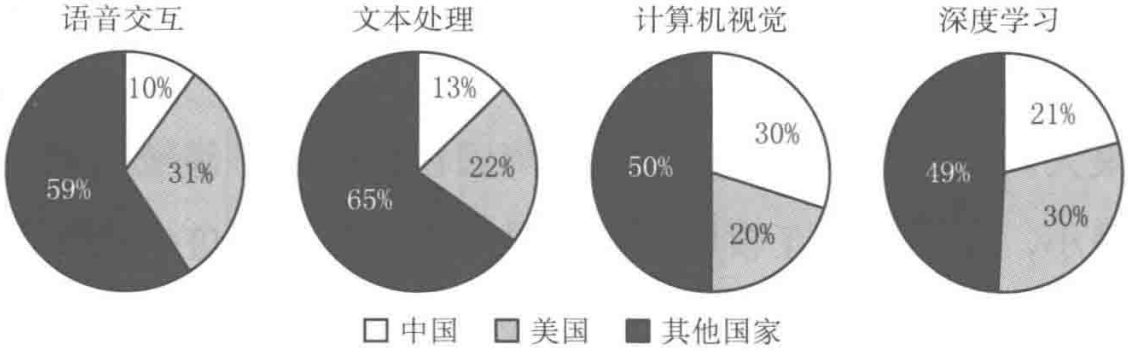
中美等国人工智能技术创新逐年发展指数

资料来源：王山、陈昌兵（2023）

在上图中，最上方带三角形的曲线是美国人工智能技术创新指数的时间序列，可以看到该指数从 2003 年就开始逐年递减，2016 年与下方的曲线相交。图中带黑色方框的

曲线为中国的人工智能技术创新指数曲线，该曲线大约也是在 2003 年左右开始持续上升，2016 年开始超越美国，跃居全球第一。

在期刊论文方面，中美两国在人工智能领域的论文总量领先全球，在四项人工智能的使能技术方面，两国的论文总量合计约占全球一半左右，见下图。



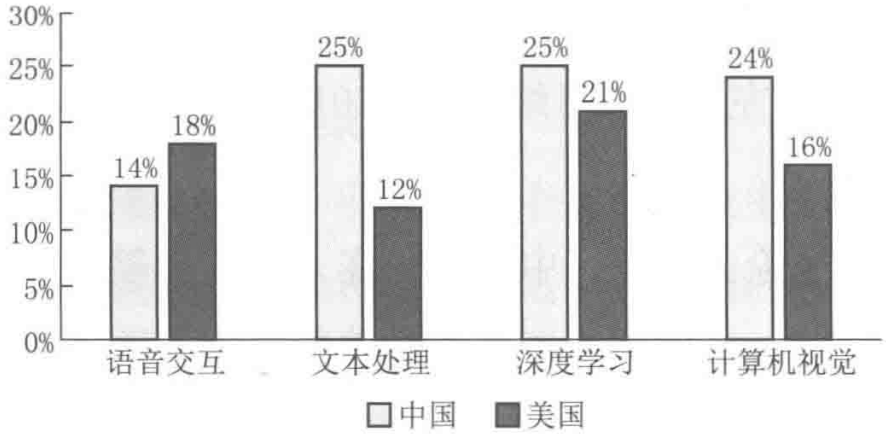
中美两国人工智能期刊论文数量占全球比重

资料来源：Web of Science，华夏幸福产业研究院

（检索时间截至 2018 年 10 月）

在期刊论文方面，中美两国的论文数量位居全球前列，且远超过其他国家。美国在语音交互、文本处理和深度学习三个方向均领先中国，尤其是在语音交互和深度学习方面的学术优势明显。在会议论文方面，中国在文本处理、深度学习和计算机视觉三个方向的论文数量超过美国，在语音交互方向上也比较接近。

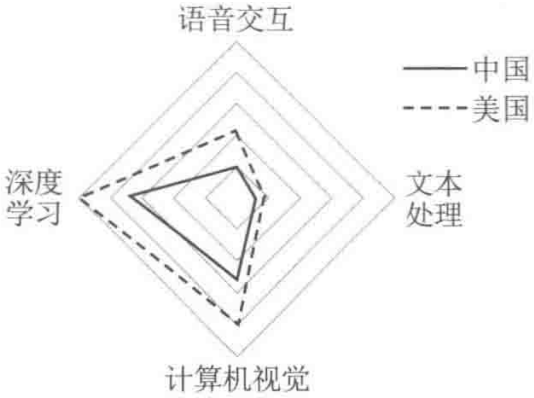
具体到在四项使能技术的细分领域，无论是基础理论还应用研究，美国均大幅领先于中国。理论研究方面，中美两国在文本处理领域的差距最小，在深度学习领域差距



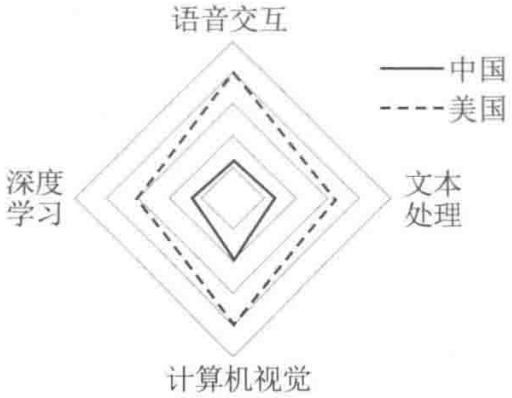
中美两国在人工智能领域的会议论文比较

资料来源：Web of Science（检索时间截至 2018 年 10 月）

最大。在应用研究方面，中美两国在深度学习领域的差距最小，在语音交互领域的差距最大。中国需要进一步加强投入和扶持力度。



中美两国人工智能技术理论研究比较



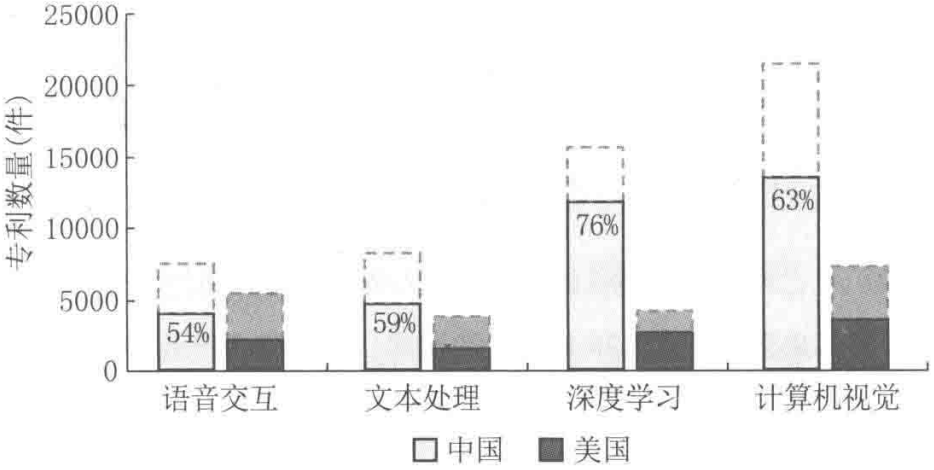
中美两国人工智能技术应用研究比较

资料来源：Web of Science（检索时间截至 2018 年 10 月）

需要关注的是在大语言模型方面，中国仍然沿着谷歌公司 Bert 模型的技术路线发展，并没有及时调整到 GPT 所采用的 Prompting 方法，美国在大语言模型方面已经拉大了领先优势。

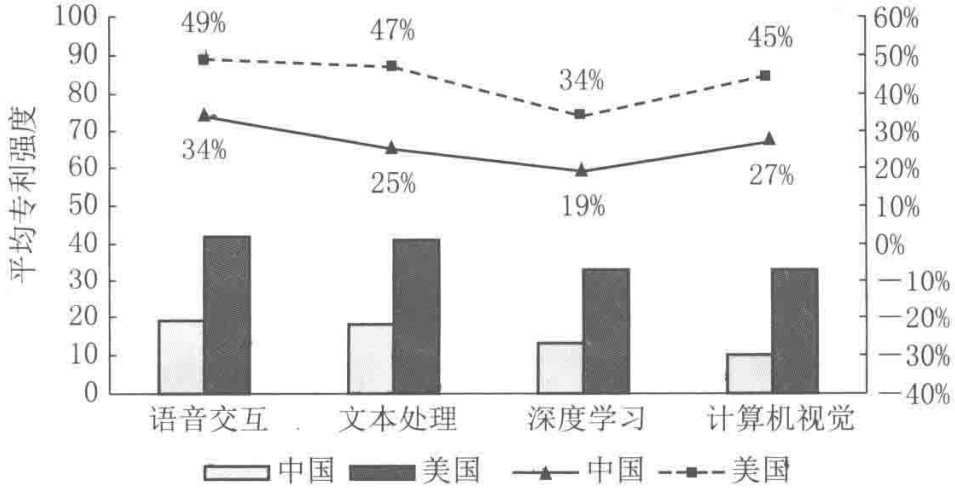
3. 应用研究的中美比较

中美两国的专利申请量均大幅领先于其他国家。其中，中国的四项使能技术专利申请量均居于首位。中国机构近三年活跃度高，超过 54% 的专利均在近三年申请。从平均专利强度来看，美国四项使能技术的平均专利强度要远高于中国，中国专利“多而不强”的现象依然存在。



中美两国专利产出规模及近三年申请比例（2018）

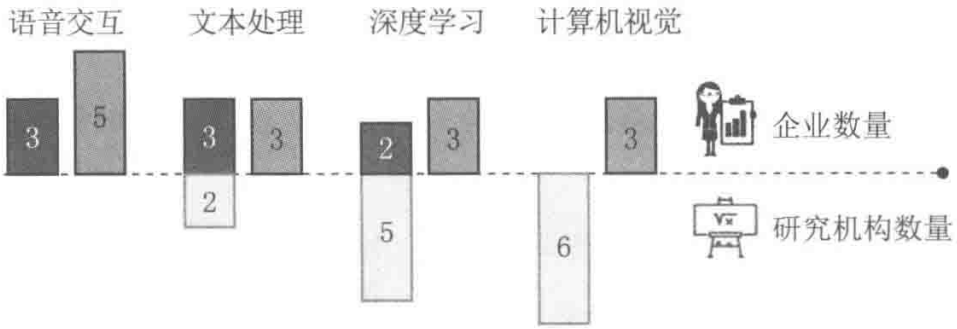
资料来源：INNOGRAPHY 数据库



中美两国人工智能平均专利强度和授权率比较

资料来源：INNOGRAPHY 数据库（检索时间截至 2018 年 10 月）

通过对前 10 位的专利权人进行分析，我们发现美国的专利权人全部为商业化企业，而中国除语音交互技术的专利权人以企业为主外，其他三项技术均存在一定比例的学术研究机构。特别是在计算机视觉领域，主要专利权人均为学术研究机构，成果的商业化将是未来关注重点。[1]



中美两国前 10 位专利权人数量分布及属性

资料来源：INNOGRAPHY 数据库，华夏幸福产业研究院
（检索时间截至 2018 年 10 月）

4. 中美人工智能的领军企业比较

从 PC 互联网到移动互联网，中国的互联网产业异军突

各阶段互联网领军企业的中美比较

资料来源：中信证券研究部（2023）

[1]《2018 人工智能产业创新评估白皮书》，华夏幸福产业研究院 2019 年。

起，借助中国庞大的市场规模和完整的产业体系，实现了后发赶超，并涌现了阿里巴巴、百度、腾讯为代表一大批互联巨头，而美国的谷歌、亚马逊、苹果、微软等公司的市值则遥遥领先。

二、各国推进人工智能的举措比较

（一）美国的人工智能进展

美国虽然在人工智能技术上处于全球领先，但是国家层面的战略一直未正式出台相关政策。在奥巴马任期的最后几个月，白宫曾经发布了三份独立报告，分别是《为人工智能的未来做准备》《国家人工智能研究与发展战略计划》和《人工智能、自动化与经济》，力图自上而下推动人工智能战略。三份报告为美国人工智能发展奠定了基础。特朗普上台之后，对奥巴马的政策进行了调整，采取了相对市场化的政策导向，但是依旧明确要求要通过政府资助研发、消除监管等措施来推动人工智能发展。

2022 年 3 月，美国参议院通过《2022 年美国竞争法》，批准 1600 亿美元用于量子、人工智能、纳米等前沿技术研究。同时，美国将中国定义成“唯一既有意图也越来越有能力重塑国际秩序的竞争对手”，围绕 5G、6G、人工智能、量子、半导体、卫星互联网、月球、深空、清洁能源等战

略领域，采取了一系列限制尖端技术和先进制造能力发展、阻断科技交流和经贸联系的做法，企图构建排华科技生态体系，加剧中美科技脱钩，使得全球科技创新生态发生重大变化。

总体而言，前期美国更多依赖企业和市场来创新人工智能技术，中美贸易战后，美国也开始更加关注举国力量的作用，开启了国家扶持人工智能竞争的进程。

（二）欧盟和英国的人工智能战略

2018 年 4 月，欧盟委员会通过了《人工智能简讯》，阐述了欧盟对人工智能的积极态度，目前欧盟委员会正在同各成员国合作，正在制定关于人工智能的协调方案。同时，欧盟决定将通过“地平线 2020”科研和创新计划框架，将对人工智能经费大幅提升 70%。

成员国中，法国推出“十五亿欧元计划”，支持人工智能研究、初创企业发展以及产业化；德国联邦政府 2018 年发布了《联邦政府人工智能战略要点》文件，要求联邦政府加大对人工智能相关重点领域的研发和创新转化的资助，并加强人工智能基础设施建设。

英国政府在 2017 年 10 月发布了《在英国发展人工智能》报告，2018 年 4 月份还发布了《人工智能行业新政》报告，涉及推动政府和公司研发、提升数字化基础设施水平、人工智能人才培养等内容。

总体而言，欧洲各国在数字经济、人工智能等领域已经全面落后，但是欧盟采取了数字化单一市场战略，同时率先占领道德高地、标准高地，相继出台了《欧盟数字贸易规则》、数字税等一系列政策和法律。欧盟的这些举措是否有利于全球经济，有利于欧洲各国发展人工智能，还需要时间的检验。

（三）日本的人工智能战略

日本政府和企业界都高度重视人工智能的发展。2017年3月，日本人工智能技术战略委员会发布《人工智能技术战略》报告，阐述了日本政府为人工智能产业化发展所制定的路线图，包括三个阶段：在各领域发展数据驱动人工智能技术应用（2020年完成一二阶段过渡）；在多领域开发人工智能技术的公共事业（2025—2030年完成二三阶段过渡）；连通各领域建立人工智能生态系统。

2022年4月，日本发布《AI战略2022》和《量子未来社会愿景（草案）》，明确在金融、医疗、运输、航空等整个社会经济体系中引进量子技术，提高生产效率和安全性。总体而言，日本在人工智能领域，无论是理论研究、智能技术应用、产业发展方面，都极具实力，尤其需要重视的是日本在机器人领域深耕多年，掌握了大部分的专利技术和关键零部件。世界各国发展智能机器人，都离不开日本的专利壁垒。这是中国发展人工智能需要关注的地方。

（四）中国的人工智能战略

中国对人工智能关注较早，中国政府很早就开始关注和支持人工智能的发展，国务院及相关主管部门相继出台多项战略规划和行动计划，鼓励和指引人工智能的发展。2015 年人工智能密切相关的智能制造被写入《中国制造 2025》，后续在《国务院关于积极推进“互联网 +”行动的指导意见》颁布，提出“人工智能作为重点布局的 11 个领域之一”，明确提到要重点推动人工智能的发展。

2016 年，在《国民经济和社会发展第十三个五年规划纲要（草案）》中提出“重点突破新兴领域人工智能技术”。2016 年 8 月，国务院发布《“十三五”国家科技创新规划》，明确人工智能作为发展新一代信息技术的主要方向。

2017 年人工智能被写入党的十九大报告，报告提出推动互联网、大数据、人工智能和实体经济深度融合，此后相关政策陆续加速出台。2017 年 7 月，国务院颁布《新一代人工智能发展规划》，从研发、工业化、人才发展、教育和职业培训、标准制定和法规、道德规范与安全等各个方面，全面规划了中国人工智能的发展战略，目标是到 2030 年使中国人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心。2017 年 12 月，工信部《促进新一代人工智能产业发展三年行动计划（2018—2020 年）》正式发布，人工智能开始加速落地，至此，中国人工智能基

本政策框架已经搭建完毕，后续则是不断细化和落实。

主要国家近年来政策规划及推动机构

国家	时间	政策及规划	推动机构
美国	2016 年 11 月	《为人工智能的未来做准备》	国家科学技术委员会、白宫科技政策办公室、国家预算办公室、人工智能特别委员会
		《国家人工智能研究与发展战略计划》	
		《人工智能、自动化及经济报告》	
	2018 年 5 月	白宫人工智能峰会	
中国	2015 年 5 月	《中国制造 2025》	国务院、科技部、工信部等
	2016 年 8 月	《“十三五”国家科技创新规划》	
	2017 年 7 月	《新一代人工智能发展规划》	
日本	2015 年 1 月	《机器人新战略》	人工智能技术战略会议等
	2017 年 3 月	《人工智能技术战略》	
印度	2018 年 6 月	《国家人工智能战略》	中央部门成立人工智能小组
欧盟	2018 年 4 月	《欧盟人工智能》	欧盟委员会等
德国	2018 年 7 月	《联邦政府人工智能战略要点》	联邦教育研究部、德国工程研究员
英国	2017 年 10 月	《在英国发展人工智能》	
	2018 年启动	《人工智能行业新政》	
韩国	2018 年 5 月	《人工智能发展战略》	科技信息通信部

资料来源：国际技术经济研究所、清华大学、平安证券研究所

2018 年 10 月 31 日，中央政治局组织第九次集体学习，学习的主题是新一代人工智能发展现状和趋势。习近平总书记在党中央十九届政治局第九次集体学习时做出重要讲话，指出：人工智能是引领这一轮科技革命和产业

变革的战略性技术，具有溢出带动性很强的“头雁”效应；加快发展新一代人工智能是我们赢得全球科技竞争主动权的重要战略抓手，是推动中国科技跨越发展、产业优化升级、生产力整体跃升的重要战略资源。习近平总书记在学习时就提出了“AI+”战略的雏形，指出要促进人工智能同一、二、三产业深度融合，以人工智能技术推动各产业变革，在中高端消费、创新引领、绿色低碳、共享经济、现代供应链、人力资本服务等领域培育新增长点、形成新动能。做出了确保这个领域“理论研究走在前面、关键核心技术占领制高点”的决策。此前的集体学习还曾就大数据战略进行过学习研讨，九次学习中就有两次与新一代信息技术相关，足见这一届政治局对该产业的高度重视。

2018 年，李克强在政府工作报告中再次论及人工智能，提出“加强新一代人工智能研发应用”。2019 年，习近平总书记主持召开中央全面深化改革委员会第七次会议并发表重要讲话，会议审议通过了《关于促进人工智能和实体经济深度融合的指导意见》。李强总理 2019 年提出“把发展人工智能作为优先战略选择。”2023 年 9 月，工业和信息化部等五部委联合印发了《元宇宙产业创新发展三年行动计划（2023—2025 年）》。

目前，人工智能正处在爆发期。中国在人工智能领域

的科学技术研究和产业发展起步稍晚，但是在国家多层次战略规划指导和政府部门的强力推动下，最近十余年的时间里抓住了机遇，人工智能的发展驶入了快车道，在全球人工智能市场中扮演了重要角色。

三、新科技革命的未来展望与伦理反思

人工智能技术的整体发展还处在初级阶段，整个人工智能行业的发展也随之面临极大的不确定性。当前人工智能领域的主要进展是基于深度神经网络的机器学习，更擅长解决的是感知问题。虽然人工智能在棋类比赛等特定任务领域已经超越人类，但在认知领域，人工智能还有较长的道路要走。

首先，ChatGPT 横空出世，给人类带来的心理冲击，远大于其实际的影响。虽然 ChatGPT 并不意味着通用人工智能的终极算法出现，但是也提醒我们人类是否在创造毁灭自己的工具。展望人工智能革命，这是人类社会普遍关注的问题。

其次，人工智能将深刻地改变就业结构。人工智能对现实世界的颠覆已在路上，人工智能带来的新挑战首先表现在就业领域。强人工智能会突破现有自动化技术的局限，推动各行各业的超常规发展，大规模的技术进步和产业升

级，将把一部分劳动者沦为结构性弃民，谷歌的专家预测这一失业比重高达98%。马克思关于“人的异化”命题，将再次引发全世界的关注和讨论。

第三，人工智能将冲击法律与社会伦理。当前比较突出的就是虚拟人问题。虚拟人犯错，责任由谁承担，损失如何控制，这一系列问题都是法律的灰色地带和空白地带。目前的智能系统已经能够自行生成虚拟人，由此带来虚拟人的言行如何规范问题。这都是以往所没有遇到的。

第四，人工智能对政府管理、经济安全和社会稳定产生深远影响。未来的通用人工智能还有待进一步完善，算法和模型（比如类脑计算）会变得更加成熟和高效。与此同时，人类也会将社会管理、政府管理的部分职责转交智能系统和智能机器来完成。当机器拥有左右人类收益甚至人类生老病死机会的权力时，整个社会的价值体系和权力结构也会随之改变。是福是祸，需要人类从长计议。

第六章

新科技革命背景下的上海科创中心建设

上海建设具有国际影响力的科技创新中心，是党中央、国务院决策部署的重大战略。习近平总书记十分关注上海科创中心的建设，提出过三个“牢牢把握”的要求，即“要牢牢把握科技进步大方向，瞄准世界科技前沿领域和顶尖水平，力争在基础科技领域有大的创新，在关键核心技术领域取得大的突破”；“要牢牢把握产业革命大趋势，围绕产业链部署创新链，把科技创新真正落到产业发展上”；“要牢牢把握集聚人才大举措，加强科研院所和高等院校创新条件建设，完善知识产权运用和保护机制，让各类人才的创新智慧竞相迸发”。

2015年以来，上海认真贯彻落实总书记“三个牢牢把握”的总要求，科技创新中心建设围绕着设计蓝图，从“提升原始创新能力、构筑创新人才高地、加快重点领域技术创新、融汇各类创新资源、提升创新治理效能”五个方面，完成了“夯实基础、攻坚突破、深化推进”三个阶段

的任务，搭建了上海科创中心的基本框架，多层次的创新平台体系基本形成，开放性的创新空间格局全面拓展，系统性的重大创新任务布局不断优化，支撑性的创新法规和政策体系日益完善，上海科创中心的核心功能加速成形。

一、上海科创中心建设成效明显

根据上海推进科技创新中心建设办公室发布的《上海科技创新中心建设报告 2022》，上海科创中心在科学发现、技术新发明、产业新方向、发展新理念等方面取得了重要进展。在科技创新的研发投入方面，上海全社会研发经费支出占全市生产总值比例提高到 4.2%，位列全国第一。人均高价值发明专利拥有量达到 40.9 件。

在创新资源的集聚方面，上海全市形成了由 100 家国家级企业技术中心、900 多家市级中心、2000 多家区级中心组成的三级网络。全国率先设立的“基础研究特区”将扩区，持续构建光子、生命科学、海洋等领域设施集群，力促 30 多家新型研发机构在上海共推产学研、竞逐高精尖。

在创新主体方面，上海已拥有高新技术企业超 2.2 万家，将培育创新型中小企业 1 万家、市级专精特新中小企业 2000 家、国家级专精特新“小巨人”企业 200 家。

在科创新赛道方面，上海正加快拓展元宇宙、绿色低

碳、智能终端、数字经济四个新赛道，将重点围绕产业、商业、生活等多个领域，推出一批融合性、综合性应用场景，推动新兴技术再升级迭代。

在科创新空间方面，上海科创中心正形成张江科学城、“大零号湾”、G60 科创走廊、杨浦科创带等东南西北四大创新基地，全市的科创空间布局基本成型。2022 年，上海工业战略性新兴产业总产值占规模以上工业总产值的比重提高到 42% 左右，有利地促进了上海经济的高质量发展。

二、新科技革命对上海科创中心建设的影响

（一）新科技革命为上海科创中心建设提供了发展机遇

随着全球化进入供应链变革的新阶段，创新发展替代成本比拼，成为区域竞争的新趋势。科技创新中心作为一个国家综合科技实力的集中体现和核心依托，其引领发展的功能不断凸显。顺应这股潮流，越来越多的国家和地区意识到新科技革命的重要意义，开始积极把握战略机遇，打造具有全球影响力的科技创新中心，以此构建区域技术进步的核心策源地，以技术进步推动产业升级和供应链重塑，进而提升国家和地区的综合实力。

1. 建设上海科创中心是国家战略的要求

随着世界经贸秩序急剧变化，供应链体系加速调整。

面对不确定的世界，新旧势力的博弈也日趋激烈，世界进入了存量博弈和增量竞赛的重大变革期。大国博弈的焦点日益转向科技创新的新赛道，以把握新一轮产业变革的主动权。

面对日渐明晰的新科技变革，众多发达国家先后提出了各自的国家战略，以提升国力、巩固地位。在众多的应对举措中，国际科技创新中心作为科技创新的主要策源地，有望成为世界创新资源的集聚中心和创新活动的控制中心，预示着一国在全球产业竞争中可能达到的最远边界和最大高度，因而成为大国竞相发展的重点。以美国为例，2011年初，美国政府就发布了《美国创新战略：确保美国经济增长与繁荣》报告，对美国未来科技发展做出了战略规划和部署。2015年进一步把美国创新战略细分为六个部分，科技创新中心是其中重要的建设内容。

科技自立是国家安全之要，产业自强是国家强盛之基。谁能抓住新一轮科技创新的机遇、抢占先机，谁就能占据主动、赢得未来。对此，中国高度重视科技创新中心的建设。党的二十大提出“完善科技创新体系”。围绕科技创新中心的建设，习近平总书记早在2013年10月十八届中共中央政治局第九次集体学习时，就提出希望中关村加快向具有全球影响力的科技创新中心进军，为全国实施创新驱动发展战略发挥示范引领作用。2014年5月，习近平总书

记在上海考察时，提出“希望上海始终立足国内、放眼全球，着力实施创新驱动发展战略，加快向具有全球影响力的科技创新中心进军”。上海国际科创中心建设不仅是中国实现创新发展的重要战略部署，更体现了习近平总书记对上海发展的殷切期望。

2. 科创中心有利于上海五个中心建设

上海已进入后工业化阶段，科技创新对可持续发展的作用日益凸显。打造具有全球影响力的科技创新中心，是上海促进经济高质量发展、应对新科技革命和新产业变革、建设全球城市的重要举措。上海五个中心建设取得了令世人瞩目的成就，但面对全球经贸格局剧烈动荡的冲击，上海的进一步高速发展也面临着一系列困难。比如，近年来上海经济的全国占比持续下降，已跌入了全国 GDP 总量的第二梯队，一度排名到全国 GDP 总量第 11 名左右。即便是在长三角区域内，无论是经济规模的体量优势，还是经济增长的速度优势，上海均大幅落后于邻近的苏浙两省。在长三角各省市 GDP 总量占比中，上海仅为 16.49%，远低于江苏的 43.76%，也大幅落后浙江的 27.53%，仅略高于安徽的 12.22%，上海的经济体量难以称得上引领区域发展。

因此，把握新科技革命的战略机遇，以科技创新重塑产业优势，提升对全产业链的控制力和影响力，进而引领长三角区域实现高质量发展，是上海迎接百年变局的应对

思路。自2010年起，上海市委、市政府就提出了“创新驱动发展，经济转型升级”的思路。新一届市委市政府提出了“深入学习贯彻党的二十大精神和习近平总书记关于科技创新的重要论述，坚持把科技创新摆在现代化建设全局的核心地位”。加快向具有全球影响力的科技创新中心进军，是习近平总书记交给上海的重大使命，也是上海五个中心建设转型升级的客观需要。

科技革命会引发产业变革，在催生出新产业的同时，也会在不同的发展阶段产生不同的主导产业。主导产业的交替升级会推动产业结构持续向高端攀升，并带动周边关联产业梯次升级，最终会拉动整个地区经济的持续增长。新科技革命催生了知识经济、网络经济和智能经济，知识创造、科技创新和企业创业能够弥补上海创新资源不足、创新企业缺乏、发展空间有限等局限，是上海提升城市竞争力的必选方向，也是推动上海五个中心建设持续更新的内在要求。

面对新科技革命的换道赶超机遇，正处于经济结构转型升级期的上海，选择了大力推进以人工智能、生物医药为代表的先导产业，在新的技术平台基础上，积极构建现代产业体系。从研发设计的产业链上游，大力发展技术密集型和资本技术密集型的产业，推动生产制造中心、金融中心、贸易中心和航运中心同步升级，夯实上海五个中心

的产业基础和微观优势，保证上海经济持续、健康、高速的发展。

在这样的思路下，上海一方面要大力推动上海科创中心建设，抓住人工智能等先导产业的发展机遇，在高端产业、高端环节上构建产业链的资源配置力、行业领导力，形成与五个中心密切关联的城市创新功能体系。另一方面，上海需要用好长三角区域一体化的市场红利，强化研发设计环节的技术实力，增强城市创新活力，整合区域市场资源，以技术进步带动产业创新，弥补上海地域空间有限的劣势，维持区域竞争的比较优势。唯此，上海才能在增强自身创新力和产业实力的同时，进一步强化对 FDI 的吸引力，强化区域新增长极的集聚效应，重塑上海国际化大都市的城市综合竞争力，实现五个中心的进化与升级。

3. 上海拥有科创中心建设的各项优势

随着经济全球化的深入发展，全球价值链日益分解和细化，组织、地域和国家对创新资源的约束日益松动，创新资源得以在全球范围内自由流动，全球创新网络因此加速形成。在这种因创新资源自由流动而形成的全球创新网络下，一些地理区位优越、产业基础较好、创新环境优良的城市更受创新资源的青睐，因此能够更广泛地集聚全球创新要素，成为全球创新网络中的节点城市，最终发展成为全球性的科创中心城市。这些科创中心城市是新科技革

命催生的区域新增长极，是推动国家和区域的创新发展的核心依托。

例如，东京集中了日本约 30% 的高等院校和 40% 的大学生，拥有全日本三分之一的研究和文化机构，以及全日本 PCT 专利产出的 50% 和世界 PCT 专利产出的 10%。硅谷以不到美国 1% 的人口创造了全美 12% 的专利产出，吸引了全美超过 40% 和全世界 14% 的风险投资。纽约集聚了全美 10% 的博士学位获得者、10% 的美国国家科学院院士以及近 40 万名科学家和工程师，每年高校毕业生人数占全国的 10% 左右。伦敦集中了英国三分之一的高等院校和科研机构，每年高校毕业学生约占全国的 40%。[1]

新科技革命加剧了创新要素的流动性和集聚性，数字经济和智能经济的网络平台模式则会进一步促进创新资源向全球少数地区或城市集聚，表现为要素资源、经济活动和科技创新的空间分布不均衡性。这种特性意味着北京、上海、武汉等节点城市更具成为网络中心节点的优势，能够凭借充足的网络通道，不断吸纳外部资源，持续对外输出影响。当国内创新城市的节点集聚力和辐射影响力超越国界时，便成为具有全球影响力的科技创新中心。这也是为什么多年以来，跨国公司落户中国时首选上海。

---

[1] 杜德斌：《对加快建成具有全球影响力科技创新中心的思考》，《红旗文稿》2015 年第 12 期。

因此，在新科技革命的背景下，上海科创中心建设需要进一步发挥好这种区位优势，需要通过强化开放性、连通性，巩固上海科创中心的网络节点优势，通过构建和强化数据资源的流通接口，让上海成为数字经济时代的新型码头，让上海成为各类科创资源集聚发展时的首选之地。

（二）上海科创中心建设需要加强创新联合体建设

新一轮科技革命是以人工智能为先导的类脑智能革命，呈现出众多新的特点：一是创新范围显著扩大，覆盖到各行各业的每一个细分领域，对人类社会的影响具有同步性、系统性和整体性特点。二是创新的深度显著加深，大到宇宙太空，小到微观粒子，对微观运行机理的研究已深入量子领域，各种新兴技术间彼此借鉴、高度融合，跨界融合的新产品、新模式带来了颠覆性变革。三是创新的速度显著加快，理论研究与应用创新的结合日趋紧密，叠加大国博弈引发的技术敏感性，需求导向的市场竞争推动创新周期不断缩短。四是创新的精度显著加强，随着人类进入数字化时代，数据成为基本的生产要素，大幅提升了近似算法和模拟现实的精度。

与新科技革命的创新特色相适应，科学研究范式也发生了巨大的转变，世界范围的学科交叉融合现象导致科技创新的难度持续上升、复杂度不断加大。仅凭个人的聪明才智和企业个体的力量，已经难以满足全产业链的广覆盖、

高技术、跨学科、迭代更新的创新需求。技术创新日益呈现出需求拉动、集群创新的特色。

对此，上海科创中心不仅需要强化国家实验室、国家重大科研攻关项目的建设，尽可能在关键瓶颈技术、重大理论研究等领域实现突破式创新，而且需要用好产业链“链主”企业和行业龙头企业的创新领导力，依托互联网平台和智能化工具，构建创新网络、创新联合体、创新产业集群，强化技术的集成式创新和跨界创新。同时，还需要进一步发挥市场在资源配置中的决定性作用，创设新型研发机构，推动互联网平台的创新分工协作，推进产学研一体化，以形成区域创新的合力，助力上海科创中心早日形成科技创新的领导力和影响力。

（三）上海科创中心需要同步推进技术创新与产业变革

科技创新是第一生产力。科技创新能够有效促进经济结构调整、加速产业转型升级，是产业变革的重要动力。随着中国经济进入提质增效的发展新阶段，人工智能、新能源、生物医药等领域的重大技术创新，日益与产业变革形成紧密关联，引发社会生产方式、生产组织模式、制造和服务模式的重大变革，甚至对传统产业产生了“归零效应”。

与此同时，新科技革命带来的重大颠覆性技术创新，诱发产业生态改变和价值链形态重构，成为新兴产业快速

崛起的必要条件。在这一颠覆式创新的过程中，面对科学创新的日趋复杂化，研发创新环节逐步模块化、标准化，以解决创新过程中的协同问题，科技研发活动因此得以独立化、产业化，并成为一种不断壮大的新型业态，创新技术本身也成为新兴产业。与此同时，科技创新加速了产业转型的步伐，促使高新技术产业迅猛发展，甚至使其成为城市的主导产业。这使得城市的功能定位和服务对象发生重大转变，当科创及其服务成为城市的主导功能之时，科技创新城市和科技创新中心也就应运而生。

面对新科技革命与产业变革的彼此促进关系，上海科创中心的建设需要牢牢把握住总书记提出的“技术创新是新兴产业发展的充分条件，是供给侧结构性改革的破题之举，竞争才能推动新兴产业发展”三条要义，从创新源头、创新转化和创新生态三个环节，推动上海科创中心持续创新，不断以新技术、新产品、新业态为传统产业赋能，推动战略性新兴产业集群式发展，以技术创新与产业变革的循环互动、迭代升级，推动上海科创中心由区域的技术转化中心升级为具有全球影响力的国际科技创新中心。

（四）上海需要同步推进虚拟场景与实体产业的创新

科技发展受多重因素影响，既有人类好奇心的驱使，也有科技发展的内在惯性，更有与经济和安全紧密相关的社会需求因素。人工智能是新一轮科技革命的主导创新，

除了作为核心的类脑智能之外，人工智能正在加速构建万物互联的元宇宙世界，加快人、机、物的三元融合，这是新一轮科技革命的鲜明特色，表现为互联网创立的虚拟空间与物理世界、实体经济的高度拟合和场景再现。

新科技革命将催生一个万物互联的元宇宙世界。所谓元宇宙世界，意指“一个平行于现实世界，又独立于现实世界的人造虚拟世界”，是一个能够“映射现实世界，又影响现实世界”的虚拟场景空间，是下一代互联网生态的重要演进方向。元宇宙的本质是全方位数字化人的感官体验——不仅是视觉、听觉，也包括触觉、味觉、嗅觉等。人们戴上耳机和目镜，找到连接终端，就能够通过创建的虚拟身份进入元宇宙的虚拟世界，宛如置身现实世界。

元宇宙强化了新一轮科技创新的广泛连接性，最终会把任何人与任何物连接起来，达成任何时间、任何地点都能建立联系的万物互联世界。这提示上海推动科技创新需要围绕场景应用，始终关注前沿领域，加速人工智能等核心技术的研发和创新，同步推进科技创新与产业变革。通过技术与产业的“四合”，即人工智能研发机构与产业集群在空间布局的耦合、元宇宙场景开发与人工智能应用的契合、网络新基建与人工智能技术创新的融合、企业主导与政府项目引导的结合，构建上海国际科创中心的核心技术优势。

三、上海科创中心建设面临的挑战

科创中心是新一轮科技革命的主要策源地，也是先进文化和先进制度的创新地，是衡量各国科技创新实力的重要指标。科创中心具有科学研究、技术创新、产业驱动和文化引领四大功能。[1]以四大功能衡量上海科创中心，在国内上海与北京同处于国内科创领域的第一梯队，相比国内其他地区在科学研究、技术实力、产业基础、企业质量等方面均拥有一定的优势，但是与硅谷、纽约等国际一流的科技创新中心相比，上海在创新能级、要素集聚、企业创新、创新机制等方面，尚有较大的追赶空间。

（一）上海需要提升科创中心的策源能级

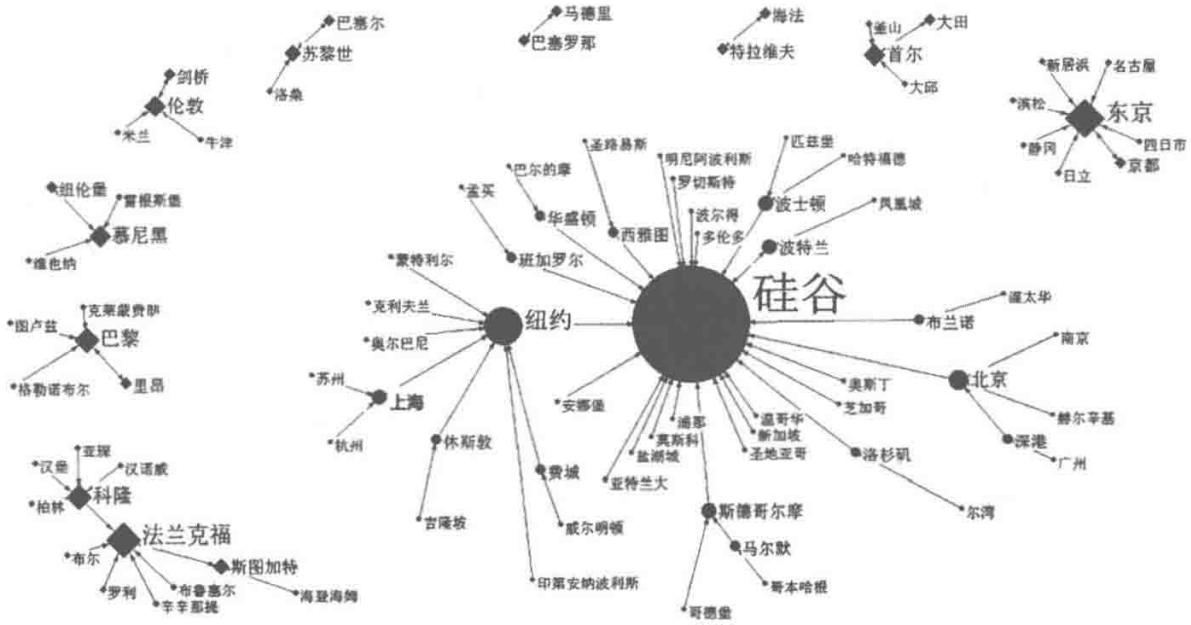
1. 上海处于全球科创网络的第三层级

在科创中心的策源能力方面，顾锋（2015）依据全球100个知名城市的研发创新合作关系，将全球的科技创新体系描绘为“1张主网、3张辅网和若干小网”，称为全球科创城市网络。[2]其中，美国硅谷是全球创新的核心策源地，是“全球科创城市网络”的超级中心，纽约在这个主网体系中位居次核心。中国北京、上海，美国波士顿、波特兰，

[1] 杜德斌：《对加快建成具有全球影响力科技创新中心的思考》，《红旗文稿》2015年第12期。

[2] 顾锋、吴文锋、曹永荣等：《上海“五个中心”新一轮发展战略研究》，《科学发展》2023年第4期。

瑞典斯德哥尔摩，则是主网体系中的第三级核心节点。与主网并列，全球科创城市的第二层级由三张“辅网”组成，分别是美国纽约、日本东京、德国法兰克福，分别在自己的周边建立了相对独立的创新城市群。比三张辅网稍逊一筹的，是法国巴黎、英国伦敦、德国慕尼黑、韩国首尔、瑞士苏黎世，同样也均在自己的周边构建了相对独立的次级创新网络，三张“辅网”和这些稍逊一筹的次级创新网络，均属于“国家级科创中心”。



全球科创城市网络

资料来源：顾锋、吴文锋、曹永荣等（2015）

上海在整个“全球科创城市网络”中处于主网体系中的第三层级，主要与硅谷、纽约等全球创新中心进行连接。换句话说，以城市间的研发创新合作关系来衡量，上海在全球创新体系中的能级比较低，主要功能是承接和转化硅

谷、纽约的科技溢出和产业项目，还缺少具有全球影响力的原创核心成果，需要努力成为全球性科技创新策源地。

2. 上海的核心科研成果增长迅猛，与第一阵营仍有距离

近年来，上海国际科创中心的建设取得了明显成效，专利发明和科学论文的数量显著提升，在国内已连续多年位列各省市创新指数排名的榜首。但是上海取得的成绩，距离具有全球影响力的科创中心尚有距离。根据世界知识产权组织发布的《2020 年全球创新指数》，全球排名前十的科技创新集群分别是“①东京—横滨；②深圳—香港—广州；③首尔；④北京；⑤圣何塞—旧金山；⑥大阪—神户—京都；⑦波士顿—剑桥；⑧纽约；⑨上海；⑩巴黎”。上海首次在全球科技城市集群榜单上跻身前十，这既反映了上海科技创新中心建设的成效显著、进步明显，也从另一个侧面表明上海的科技创新能力仍需要提升。而根据世界知识产权组织最新的《2022 年全球创新指数（GII）》报告，上海—苏州创新集群则排名全球第 6 位，超过日本的大阪—神户—京都创新集群。

在世界知识产权组织的这份科技城市集群的排名中，其排名依据主要有两个：一是《专利合作条约》（PCT）所列的专利申请量；二是科学引文索引扩展版网站（SCIE）公布的科学出版物数量。其中，上海的 PCT 专利申请量为 13347 件，远远落后排名第一的东京—横滨都市圈 113244

全球前 20 位科技创新集群排名

集群排名	集群名称	经济体	PCT 申请量	科学出版物	在 PCT 申请总量中的份额（%）	在出版物总量中的份额（%）	共计	2015—2019 年排名
1	东京—横滨	日本	122526	112890	10.7	1.6	12.3	1
2	深圳—香港—广州	中国 / 香港	94340	133327	8.2	1.9	10.1	2
3	北京	中国	32016	260937	2.8	3.7	6.5	3
4	首尔	韩国	46273	124530	4	1.8	5.8	4
5	加利福尼亚州圣何塞—旧金山	美国	42884	58087	3.7	0.8	4.6	5
6	上海—苏州	中国	22869	148203	2	2.1	4.1	7
7	大阪—神户—京都	日本	34738	50605	3	0.7	3.8	6
8	马萨诸塞州波士顿—剑桥	美国	16172	73457	1.4	1	2.4	8
9	纽约州纽约市	美国	13020	73623	1.1	1	2.2	9
10	巴黎	法国	14147	62793	1.2	0.9	2.1	10
11	加利福尼亚州圣地亚哥	美国	19363	20688	1.7	0.3	2	11

续表

集群排名	集群名称	经济体	PCT申请量	科学出版物	在PCT申请总量中的份额（%）	在出版物总量中的份额（%）	共计	2015—2019年排名
12	名古屋	日本	18623	17261	1.6	0.2	1.9	12
13	南京	中国	3620	103260	0.3	1.5	1.8	15
14	杭州	中国	8568	55312	0.7	0.8	1.5	18
15	加利福尼亚州洛杉矶	美国	10515	43172	0.9	0.6	1.5	14
16	武汉	中国	4126	80002	0.4	1.1	1.5	20
17	华盛顿哥伦比亚特区—马里兰州巴尔的摩	美国	4727	75104	0.4	1.1	1.5	13
18	华盛顿州西雅图	美国	11943	20396	1	0.3	1.3	16
19	伦敦	英国	4936	56911	0.4	0.8	1.2	19
20	大田	韩国	10286	23812	0.9	0.3	1.2	22

资料来源：世界知识产权组织（WIPO）《2022年全球创新指数（GII）》报告

件，仅为第一名的八分之一。同时，上海的科学论文数量远远高于专利申请数量，这既表明了上海强大的学术科研能力，也反映出企业作为创新主体的引擎作用还需要加强，因为 PCT 专利主要面向全球，服务于企业的国际化战略和国际市场。

3. 上海有望在人工智能等领域实现赶超

洞悉全球科创中心竞争态势的同时，上海也在积极打造科创中心的核心竞争力。近年来，集成电路、人工智能、生物医药三大先导产业得到了重点发展。以人工智能为例，上海在核心技术、场景运用、生态建设、人力资源等方面持续发力，世界级的人工智能产业集群初具规模。围绕“东西集聚、多点联动”的布局，已经规划建成浦东张江、徐汇西岸、临港新片区、闵行马桥四大人工智能产业承载区，市北高新、长阳创谷、虹桥智谷等特色园区加快建设，华东无人机基地等区域产业正集聚发展。上海量子科学院研究中心、上海脑科学与类脑研究中心张江实验室、上海数据交易所、白玉兰开源开放研究院等一大批新型机构建成运营。

目前，上海科创中心已经在类脑研究、计算机视觉、核心芯片、算法平台、智能机器人等人工智能关键基础设施领域具备领先优势。比如，在核心芯片方面，依图科技的“求索”系列云端芯片、天数智芯的 7 nm GPU 芯片、华

为全栈式人工智能战略和人工智能芯片、寒武纪云端和终端芯片等一批智能芯片流片量产。上海已基本形成人工智能全栈技术产业链，逐步从低阶感知智能向高阶认知智能、决策智能跃迁，以超大规模预训练模型为代表的核心算法将成为解决核心问题的关键所在。

在算法平台方面，上海人工智能实验室、上海脑科学与类脑研究中心、上海人工智能算法研究院、上海自主智能无人系统科学中心、上海期智研究院、上海人工智能研究院、中国科学院计算技术研究所智能处理器研究中心、微软亚洲研究院（上海）等一批基础研究和前沿理论研究机构落户上海。在应用转化平台方面，上海白玉兰开源开放研究院、微软—仪电人工智能创新院等技术开发转化平台，主要从事关键共性技术的转化落地。商汤科技、依图、上汽、腾讯、深兰、寒武纪、优刻得、亚马逊、IBM、阿里、百度、华为、科大讯飞等头部企业均在上海设立了人工智能创新中心，人工智能的产业生态体系业已成型，上海有望引领全球人工智能行业的新一代浪潮，掀起智能产业变革的蝶变。

（二）上海需要大力提升产业社区的要素整合能力

在创新要素的整合方面，上海的创新要素集聚程度居全国前列。根据《中国区域科技创新评价报告 2022》，上海、北京、天津、广东、江苏和浙江综合科技创新水平指

数值高于全国平均水平的地区，位列全国第一梯队。其中，上海多年来综合科技创新水平的排名一直是全国第一，科技活动财力投入指数排名也是全国第一。2021 年上海 R&D 经费支出 1615.7 亿元，占 GDP 比重 4.17%，位居全国第二。长三角地区在上海的引领下，已经成为国内最具竞争力的科创共同体。然而，对标建成具有全球影响力的科创中心，上海在创新要素的资源整合方面尚需努力。

1. 创新主体资源需要加强关联

习近平总书记在党的二十大报告中指出："加强企业主导的产学研深度融合，强化目标导向，提高科技成果转化和产业化水平。强化企业科技创新主体地位，发挥科技型骨干企业引领支撑作用，营造有利于科技型中小微企业成长的良好环境，推动创新链产业链资金链人才链深度融合。"总书记的这些重要论述，明确了企业作为创新主体重要地位在中国科学院第二十次院士大会上，习近平总书记强调："要发挥企业出题者作用，推进重点项目协同和研发活动一体化，加快构建龙头企业牵头、高校院所支撑、各创新主体相互协同的创新联合体，发展高效强大的共性技术供给体系，提高科技成果转移转化成效。"

上海科创中心未来的发展重点，需要深入把握科技创新的规律，加快构建创新联合体等新型科创机构，将各创新主体有机连接在一起，打破部门局限、条线制约、行业壁垒，

聚合创新要素资源，构建目标一致、利益共同的创新联合体。

2. 创新平台载体需要强化资源聚合功能

上海市政府于2022年印发了《上海打造未来产业创新高地发展壮大未来产业集群行动方案》，提出“建设核心技术自主创新的未来高地，成立5家左右未来技术学院，培育15个左右未来产业创新中心，建设一批创新联合体，打通基础研究、应用基础研究到产业化的双向通道。做强未来产业集群发展的未来引擎。打造5个未来产业集群，建设15个左右未来产业先导区……营造要素集聚、开放包容的未来生态”的发展目标。

这一系列的重大举措为上海打造全球的创新高地提供了坚实的支撑，未来需要以“行动方案”为蓝本，进一步发挥网络平台的聚合效应，以包容的创新氛围、便捷的创新服务、公允的成果评估体系、畅通的转化渠道，提升上海科创中心的吸引力，以聚集更多的一流企业、一流人才、年轻学子，形成比肩硅谷的全球创新中心。

3. 科技园区需要向产业社区转型升级

城市近郊科技园区如紫竹科技园区、浦东张江、徐汇西岸、临港新片区、闵行马桥，其中既有新兴的科技园区，也有发展数十年的老牌园区，均已聚集了大量的人工智能、电子信息、高端制造等战略性新兴产业和新兴先导产业，呈现了欣欣向荣的发展局面。未来这些科技园区需要针对

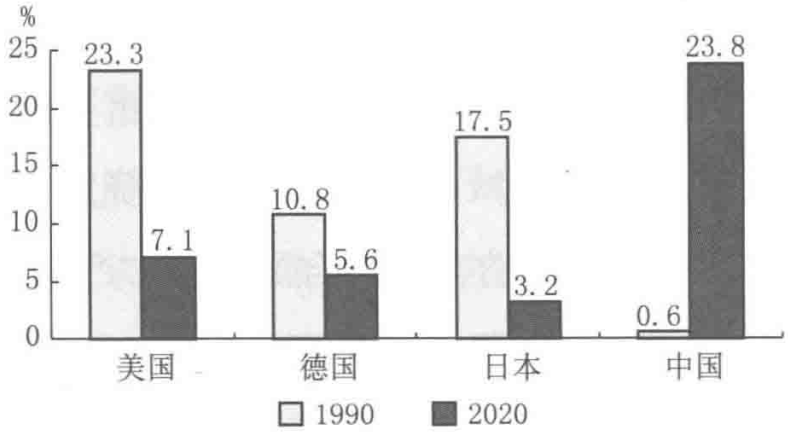
企业员工和客户普遍存在的商业、医疗、教育等公共服务需求，以及舒适住宅和社区环境的居住需求，便捷出行的交通需求，重点提升产城融合度。可以围绕产业功能、创新功能和生活功能，打造功能融合的未来“产业社区”。通过建设公共空间完善、服务设施配套、二三产业融合、智能功能健全、环境景观宜人的新型产业社区，缔造未来城市的创业、创新、休闲新空间。

（三）上海需要积极提升本地企业的创新能力

1. 上海的发展过于依赖外资和国企

在创新能力方面，企业作为科技创新的重要载体，也是实施产业升级的主体，企业的创新动能和创新成果，大致能够衡量一个城市（地区）的科技创新能力。一是上海的企业构成为国资和外企为主，是偏外向型经济结构。对外贸易在上海的经济结构中占比一直较高，外资在其中发挥了巨大作用。二是上海高新技术出口占比持续增长。2021年上海高新技术产品出口占全市比重为38.5%，超过全国29.11%的平均水平。上海对外贸易的靓丽成绩得益于中国制造的崛起，自2005年起中国就已经取代美国成为全球高技术产品出口的冠军，2020年中国高技术产品出口占据了全球23.8%的份额，远超美国16.7个百分点。

然而，在这份靓丽的外贸成绩单背后，外资却发挥了巨大的作用，外商投资企业创造上海超过四分之一的GDP、



各国高科技产品出口占世界市场比重（1990 年与 2020 年）

资料来源：德国联邦外贸与投资署（GTAI）

三分之一的税收、三分之二左右的货物进出口工业总产值，以及五分之一的就业岗位，占据了上海高科技产品对外出口的较大比重。2021 年上海货物出口总额为 15718.67 亿元，其中外商投资企业出口 9111.30 亿元，占比 57.96%。2022 年，上海出台多项惠企便利措施，推动外贸进一步增长，其中外资企业进出口 2.56 万亿元，增长 2.2%，占全市进出口值的 61.1%；民营企业进出口 1.23 万亿元，增长 11.9%，在全市进出口值中的占比提升 2.3 个百分点至 29.4%。

2. 外资的技术溢出效应有限

上海科技创新需要聚集各类创新资源，外资是其中不可或缺的重要力量。然而，需要引起高度重视的是外资的不确定性和高度流动性，上海科创中心需要尽可能提升创新主体的相对稳定性。近年来，由于国内成本上升、美国持续打压等原因，叠加新冠疫情和国际局势动荡的影响，新晋外资的落地和在沪外企追加投资（包括新设绿地投资

和并购投资）的意愿减弱，部分在沪外资呈现加速流出迹象，稳外资、稳预期成为近年来各级政府的一项重要工作。在关注外资外贸的同时，也提醒我们拥有强大的本土企业，才能真正建立一个相对更稳定的企业群体，同时也能提升上海对外资的吸引力和竞争力。

相关研究表明：多年以来，外资对上海经济增长的技术溢出效应较低，远低于港澳台企业和内资企业，与其进出口规模和对外进出口占比形成鲜明的反差。同时，外资也在其所在的行业内普遍构建了较高的专利壁垒，对后来的竞争者形成压制和阻碍。与促进行业发展相比，外资更看重的是中国的市场和企业的利润，对行业的技术溢出反而是高度防范的。因此，我们在推进上海科创中心的建设时，一方面要清醒地认识到外资对上海高科技产品出口的贡献，尤其是外资对促进上海技术进步和产业发展的巨大作用，比如上海引进特斯拉就成功地拉动了国产电动车行业的市场竞争，推动了电动车行业的技术进步。但另一方面，我们也要清醒地认识到，核心技术是买不来的，外资外企是否升级产品、升级技术，取决于中国的本土企业的技术进步程度和创新速度，优惠的招商引资政策只是备选因素。加速上海国际科创中心建设，必须立足本土、以我为主，扎实推进基础研究、技术应用，同时坚持开放发展、合作交流的方针，用好国内、国际两个市场的资源，尽早

培育和提升本土企业的科技策源能力。

3. 上海需要认真思考“谁来创新”的问题

从创新主体来看，创新的主体有外资企业、央企、地方国企、民营企业、高校和科研院所、个人，其中企业创新的主体是外资、国企和民企三类。目前，外资和国企在上海的经济结构中占比较高，前者没有主动转让技术的动机，后者则过于求稳、创新乏力。

因此，国企的创新更大程度上是技术发明，发明成果搞出来，公开向社会展示之后，多数会被束之高阁，而不会真正投入批量使用。除了技术可靠性、技术稳定性等关键因素以外，风险与成本的理性选择，也是重要的影响因素。技术创新有六个阶段：创新构思阶段、项目确定阶段、研究开发阶段、工业设计与产品试制阶段、规模投入与批量生产阶段、技术实现阶段。其中，工业设计与产品试制阶段通常是对原有设备和工艺的技术改进，创新难度低、规模小，但工艺设计创新带来的利润收益和宣传效果可观，所以也是最能体现创新成绩的阶段。而进一步进入规模投入与批量化生产阶段，由于技术风险几乎为零，但经营风险剧增，新产品、新工艺没有达到经济批量之前，都会受制于固定成本的制约，往往很难带来经济效益，更难形成经济优势。[1]这不符合国有企业求稳、求平衡的特性。所以

[1] 陆卫国、郑毅、曾繁荣：《中小企业技术创新不同生命周期阶段与资金的需求》，《社会科学家》2005年第3期。

多数国企的创新往往止步于技术发明，止步于新产品试制阶段，而少有大规模投入与批量化生产。

（四）全球科创城市的竞争已全面展开

1. 全球科创中心城市的竞争早已开展

在引领区域创新方面，纽约、伦敦、新加坡、东京、首尔等先后提出了建设全球或区域科技创新中心的目标，并出台了相应的战略规划。英国卡梅伦政府于2010年就推出了“英国科技城”战略，提出打造东伦敦地区世界一流的国际技术中心的设想。从2011年正式启动以来，凭借“知识（服务）+创意（文化）+市场（枢纽）”模式，吸引了数万家高科技公司进驻科技城，目前成为堪比硅谷的国际科技创业中心，是世界第三大科创产业聚集地。

2008年金融危机后，美国也意识到了产业空心化的危害，试图通过再工业化战略，借助新科技革命引导产业回流，在新的技术平台上重构全球分工体系。[1] 2002年，纽约宣布要将城市打造成“创新之都”和美国“东部硅谷”。2010年，纽约市政府又率先提出建设“全球科技创新中心”的目标，这也是全球首次提出“科技创新中心”的概念。经过十余年的发展，纽约的创新生态系统已经形成，截至2021年，纽约科技行业的从业人数达到了37.2万人，占全

[1] 郭进、杨建文：《美国再工业化战略对中国产业发展的影响及对策》，《经济问题探索》2014年第4期。

市劳动力比重高达 12.6%，经济总量超过 1247 亿美元。

日本地域狭小、资源匮乏、市场容量有限，所以多年来一直高度重视科技创新。早在 2014 年，日本就提出打造“全球领先的创新中心”的宏伟战略，明确了东京成为全球创新网络枢纽的发展目标。东京科创中心属于典型的政府主导型模式，是“工业（集群）+ 研发（基地）+ 政府（立法）”的创新模式，是产业集群与国际枢纽的高度融合，目前东京已经成为集制造业基地、金融中心、信息中心、航运中心、科研和文化教育中心及人才高地于一体的科技创新中心。

2. 国内科技创新中心的城市竞争更是遍地开花

全国已形成四大创新集聚区，分别是以北京为中心的京津冀创新集聚区，以上海为中心的长三角创新集聚区，以广东为中心的珠三角创新集聚区，以成都、重庆、武汉、西安为中心的区域性创新集聚区，点、线、面梯次联动的区域创新格局正加快成型。[1] 根据各省市“十四五”规划，目前已有北京、上海、广东、重庆、湖北、江苏、浙江、河北、广西、贵州、江西等十余个省区市在“十四五”规划中提出了加快打造（区域）科技创新中心。南京、杭州、成都、重庆、济南、青岛、西安、沈阳、兰州等城市，也

[1] 钟源:《多地布局“十四五”区域创新中心建设》,《经济参考报》2021 年 11 月 29 日。

均在“十四五”规划中提出创建综合性国家科学中心或科技创新中心。其中，截至2022年7月，国务院审核同意的全国科技创新中心有北京、上海、粤港澳、成渝和武汉五个。在细分领域，科技部已批复建设的国家技术创新中心有19家，工业和信息化部正式公布的国家先进制造业集群有45个。

3. 上海科创中心建设需要实施差异化战略

各地积极创建国家科技创新中心，意图都是抢抓国家战略科技力量布局的先机，以技术创新带动产业兴旺和区域创新，重塑区域经济发展的增长引擎。因此，此轮国内展开的科技创新中心竞争，将决定未来中国的区域经济格局，上海必须直面区域科创竞争的压力。各地高度重视科技创新中心建设，也提醒上海需要找准自身优势，遵循要素集聚、创新集聚的规律，注重体现地方特色、发挥比较优势，以产业链、价值链和创新链为依托，扬长避短、突出创新的重点，错位协同发展，带动和引领全国科技创新实力提升。北京、上海、深圳等城市应瞄准全球科技创新网络的关键节点和全球城市体系的金字塔尖，打造全球性科技创新中心；武汉、西安、南京等城市应力争成为亚洲具有广泛区域影响力的科技创新中心。

如果说北京作为国家级科创中心，目前正在积极追赶全球第二梯队的法兰克福、东京，那么国际专利申请数量

仅为北京二分之一的上海，面对国内如火如荼的城市科创中心竞争，则需要摆脱城市地域空间有限、商务成本高企、企业数量有限、研发成果难以本地产业化等局限，用好长三角一体化的政策红利，构建区域集群创新的新模式。以上海为“科技创新 + 市场创新”的龙头，带动江苏、浙江、安徽的“技术发明 + 企业创新”，以区域一体化创新整合长三角的创新资源、产业资本和区域市场，构建长三角创新产业集群，形成长三角全球科技创新中心。

（五）上海需要积极推进创新的体制机制改革

在内外部创新环境方面，科技创新需要突破体制机制的制约，这一观点已得到广泛的共识。但是改什么，向什么方向改，如何改，却众口纷纭，莫衷一是。这里我们不讨论具体的建议，只是回溯到西方为何能开启科技革命的本源，围绕知识产权制度这个核心因素，以问题为导向来讨论科技创新中心建设普遍面临的体制机制问题。

知识产权制度是近现代西方崛起的核心密码，是走出中世纪的关键因素之一。中世纪的欧洲有多数地区实行长子继承制，这一制度保证了封建制度的稳定，但也迫使无继承权的次子们不得不离开乡村另谋生路，这促进了近代欧洲小城镇的兴起。中世纪的欧洲城镇有两个重要功能：一是“城”的防御功能；二是“市”的交易功能。失去土地的农民和次子们为了在城镇谋生，要么出卖体力参与城

镇的军事防御，要么出卖体力和智力参与城镇的商品交易，这与现代社会的农民进城存在共同之处。因此，进城的农民和城市居民在城镇商品经济的分工模式下，就格外重视赖以谋生的技能，并将之人格化为专利，由此发展出现代的知识产权制度。这就是为什么西方格外重视创新的原因，因为创新所形成的技能具有人格属性，能够谋生、能够交易、能够创富。有了知识产权等法律制度的保证，人格化的创新不会被轻易剥夺，而是能够被合理计价，顺利进行交易。

反思我国的科技创新，无论是举国体制，还是大众创新，始终面临着“创新动力哪里来”的问题。因此，体制机制的改革首先要回答“动力何来”的问题。如果能够真正做到谁发明、谁受益，能够顺利让创新者从创新中直接受益，将创新顺利地转化为真金白银的收入，创新的动力自然会被激发，上海科创中心的建设也必将迎来更多的参与者和创造者。

四、上海科创中心如何用好新科技革命的机遇

（一）上海科创中心建设要立足现实、长远谋划

1. 上海科创中心需要实施差异化的发展战略

任何科技创新中心的兴起都是时间因素和地理空间交

互作用的结果。迄今人类历史上共发生过五次创新资源跨国大转移，每次转移都会在新的区位造就新的科技创新中心，并推动大国实力的此消彼长，导致全球霸权的更迭。上海科创中心的建设同样需要天时、地利与人和的共振，不仅需要等待时机，顺势而为，而且更需要长期努力，持之以恒。

上海科技创新中心建设注定是一个需要长期努力的过程。当前，国内科创中心的竞赛已经展开，对比北京、深圳的科创中心建设规划，各地都在瞄准全球科技创新网络的关键节点，突出自身既有的优势，实施差异化的发展战略，力图构建区域特色的科技创新中心。比如，在人工智能领域，北京、上海、深圳都制定了符合自身的产业层次的方案和举措：深圳产业基础雄厚，制造业非常发达，年轻人口居多，因此，深圳人工智能战略的重点是赋能千行百业，是改造传统产业与发展新兴产业并重的思路。北京有很强的科研优势，有中央部委所在地的政治优势，同时也有中关村的平台优势。因此，北京聚焦国家战略，优先解决技术卡脖子的瓶颈问题，重点发展芯片制造等产业。上海的优势在于市场化的氛围很好，城市的开放度、包容度均领先全国。同时，上海国企和政府的力量强大，而短板则是民营企业。因此，上海的发展策略应当是继续做强优势领域，积极弥补民营企业的短板。制定差异化的长期

规划，重点突出市场竞争的制度激励。

2. 上海科创中心需要坚持排头兵的功能定位

全球科技创新中心是指能够代表世界科技发展前沿和最先进生产力的国家或地区。进一步细分，科创中心有科学中心、技术中心还是产业中心之分，上海需要回答“发展的新赛道在哪？战略性新兴产业选择是否正确”的问题。

中央赋予上海的战略使命是改革开放的排头兵，科技创新的策源地。这提示上海科创中心的建设需要形成强大的国际竞争力和影响力，要成为全球新知识、新技术和新产品的创新策源地和全球生产制造基地。因此，上海科创中心的功能定位就必然是全面推进，不可偏废。即不但需要大力发展基础理论研究的科学中心，而且需要积极推进应用技术研发与市场模式的创新，同时还需要大力发展战略性新兴产业和人工智能等新兴行业。三者之间应当形成彼此促进、相互支撑的关系，这也是上海科创中心未来需要重点探索和创新的地方。

3. 上海科创中心需要坚持开放发展的方向

随着全球创新网络的日趋成型，科技创新已不再是单打独斗的个体行为，不仅需要良好的地方创新能力，更需要融入全球创新网络，形成团队创新模式和组织创新联盟。

为此，上海建设科创中心建设，一是要把握全球科技创新的趋势，主动对接创新发展国家战略，构建上海科创

中心与国家科创战略和发展规划的紧密关联，力争能够及时、充足地得到国家政策、资金和项目的支持。

二是上海科创中心建设要与全球知识创新紧密结合。在立足自身优势和特色，制定科技创新中心城市的长远战略规划和近期目标的同时，上海也需要把科技研发作为城市经济的主导产业，以开放的姿态拥抱全球的各类知识，以世界一流大学和创新企业的知识创新，推动上海建设全球知识的集聚地和转化地。换句话说，上海要成为新知识、新技术的“十里洋场”，成为研发活动集中的地方，以此强化全球创新网络的中心节点地位。

三是上海要立志成为全球科技创新企业和科技研发机构的集聚高地。全球性的创新机构和研发组织集聚，不仅能够开阔上海的眼界，激发创新的思维，同时也会带来资源集聚效应、人才集聚效应和经济增长效应，有利于促进上海融入全球创新网络。以机构和企业集聚，带动上海的开放发展，将是上海科创中心建设的重要举措。

（二）推进上海科创中心建设要用好新型研发机构

1. 新型研发机构要能体现人才的价值

目前，国内的产学研创新主体大多各成体系，缺少创新的整体关联。高校、研究机构、企业与科技园区等创新主体，各有各的优势，各有各的领域，但多数都是自成体系，缺少合作与互动，更缺少对创新空间的统筹布局和整

体安排。中心城区的大学、园区和社区等创新资源虽然集中，但彼此之间也欠缺联动机制，未能形成创新集群的集聚效应和产业联动效应，人才的创新价值也无从体现。对此，上海需要发挥大学人才集聚、科研机构众多的优势，以张江科学城、大零号湾科创基地等新型平台为依托，积极发展各类新型研发机构，创新的重点是让人才有创新的舞台，技术有应用的对象，科学家能够通过创新获得财富，人才的价值能够充分展现。

2. 科创中心要集聚各方人才，建设世界级人才高地

科技创新中心建设需要庞大的人才队伍支撑，鉴于“核心技术买不来”的现实，迎接新科技革命的人才队伍也要坚持两条腿走路，自主培养与全球引进应同步实施。一方面，强化高等院校、科研院所对基础理论研究，同步推进新科技革命相关的人工智能、生物医药、航空航天等专业的人才培养，以技术攻关和技术创新带动人才培养，为新科技革命准备庞大的技术人才梯队。另一方面，领军人才对技术和行业的发展至关重要，面对西方的创新优势，上海科创中心建设需要敞开大门，积极引进全球优秀人才、创新企业。

在具体举措上，一是可以进一步加大改革开放的步伐，以合作办学的形式，引进海外一流大学，促进国内高等教育体系改革，提升人才队伍培养能力。二是加强智慧城市

建设，提升智慧城市的生活便利度，提升城市的人才吸引力、创业便利度等。三是进一步降低人才引进的门槛，比如永久居留权门槛、放宽签证期限等政策，以及对标 TPP 等自由贸易区的政策，对外籍高层次人才工作许可实行负面清单管理。四是优化引才的服务环境，比如继续对引进人才提供配套的就业、子女教育、生活补贴、购房优惠等政策，为创业创新人才提供一个舒心的生活环境。五是实施更加便捷的出入境管理政策。如增加免签证国家、延长免签停留时间、通关便利化、单一窗口等。

（三）上海科创中心需要大力发展民营经济

激发创新动力，除了解决微观主体的创新利益分配问题外，还需要从更高层次回答好“谁来创新”的问题。各类主体如何参与创新，涉及更深层次的利益分配问题，是中国当前改革的深水区。如前所述，外资和国企在上海的经济结构中占据了极大比重，民企的发展相对薄弱。三类企业里，外资有技术优势、产品优势和研发优势，但技术转让和技术合作的意愿不高，甚至设置专利壁垒，试图以此巩固优势地位。而国企更偏重的稳定经营、技术发明，真正能够规模化应用的创新较少。相比国企，民企出身草根，背景和实力都有限，创新谋生的动力也就更为强烈。如果能够将其对新技术的热情与高校和科研院所的研发能力结合，无疑能够创造出巨大的活力。事实上，近年来我

们看到越来越多的民营企业家主动走进校园和科研院所，以项目需求带动应用研发的模式，合作开发了许多创新产品。比如智能呼吸诊断分析、智能空气净化等领域，就有许多这样的民营企业，都是高校和科研院所的科研项目开花的结果。

在这一方面，对比北京和深圳两个科创中心，上海尤其需要加大马力。如果说北京是国家级的科技创新中心，能够更方便地获取国家部委的项目支持，通过中关村将国家的重大科技攻关成果转化为产业应用，更能代表国家创新的业绩；那么深圳的崛起主要得益于发达的民营经济和外向型的制造业。20 世纪 80 年代，深圳起步建设之初，主要得益于官员下海和全国支援。这些下海官员是当时国内最具国际视野的企业家，敢于主动放弃条件优越的“铁饭碗”，因而创新创业的动力十足，有“置之死地而后生”的勇气和能力，这也造就了当今深圳创新的特殊基因和创新模式。

与国内两大科创中心相比，上海的海派底蕴更为浓厚，具备大力发展民营创新机构的可能。上海的民营机构与北京和深圳相比，除了自主研发的团队和本土型跨国公司之外，可能海外人才回国创业的比重会更多一些。北京科创中心是国家创新技术的转化落地，深圳科创中心是国资背景的人才与潮汕文化的探索精神相结合，而上海的科创中

心则需要弘扬海派文化的包容精神，将更多的国外先进技术引进到我们的自主创新体系里来。上海民营科创企业的这一作用，在美国大搞对华脱钩的战略下，是突破“小院高墙”封锁的重要途径。

（四）大力推动创新体系的体制机制改革

创新体系的体制机制改革有许多领域，当下宜重点选择社会普遍关心的“政府职能改革、企业创新主体、金融扶持创新”三个领域，分别对应浦东新区社会主义现代化建设引领区、上海自由贸易（试验）区、长三角一体化示范区等三个国家级的改革开放战略，以便推出重大创新举措，用好先试先行的政策红利。

1. 进一步推动政府职能改革

在经济发展中，政府与市场存在某种跷跷板关系，政府过于强大，市场必然受到抑制；政府职能不转变，市场机制就起不来。为了鼓励创新，现阶段政府必须拿出自我革命的精神，坚决转变职能、创新管理。理顺统筹创新的体制机制，需要处理好三种关系：

一是正确处理好科学研究与创新发展的关系，既要完善政府的投入机制，保障科学研究的深入开展，尤其是要进一步加大对基础研究的财政投入力度，弥补中国长期对基础理论研究的忽视。同时，也要正确处理好创新发展的市场竞争机制问题。政府不能大包大揽，需要营造良好市

场竞争环境，更多地以政府引导基金、政策性金融工具等形式，引导和鼓励市场化的应用创新。

二是正确处理科技创新与经济发展的关系。科技创新是第一生产力，举国体制突破关键性的瓶颈技术后，政府发展经济面临着如何推广创新成果、实现成果转化的抉择。可以借鉴特斯拉公开汽车源代码的模式，让市场主体共享基础创新的成果，并推动应用转化。通过让利社会带动整个行业发展，进而实现经济的增长。

三是正确处理国有企业与民营企业的关系。政府需要营造公平竞争的环境，更多通过一视同仁的市场规则维护市场秩序，尽可能减少对特定行业、特定地区、特定企业的优惠政策。当技术创新替代垄断资源，成为企业成长的主要途径，市场主体自然会争相创新发展。

2. 强化企业作为创新主体的地位

企业是市场经营的主体，也是经济发展的重要力量，在国家科技创新体系中占有十分重要的地位。强化企业的创新主体地位，关键是如何去做。

当前，一是要针对企业创新面临的原创力、组织力、控制力不足等问题，展开针对性调研，重点分析企业家创新的障碍、企业创新活力缺失的原因、如何发挥企业家作用等问题，为制定一揽子政策提供第一手的调研信息，做到有的放矢。

二是要积极鼓励企业参与国家和上海市的重大科技攻关计划。只有企业有机会与国家部委、各级政府、科研部门在实际工作中形成密切接触，企业才有可能与国家和社会建立起广泛的联系，才能更好地形成社会认同的创新共同体。

三是建立企业主导的联合攻关机制，引导企业加强创新研究。研究和应用脱节的问题由来已久，国外对这一问题的解决办法是让企业来主导科技创新和应用开发。上海可以学习借鉴发达国家的经验，加以改造后为我所用。

四是用好浦东社会主义现代化建设引领区的政策，围绕企业创新能力提升、创新平台建设、科技成果转化、科技人才发展、科技金融融合、创新创业载体等科技创新活动全要素、全过程，出台以企业创新为主体的项目支持、贷款贴息、担保贴费、后补助、股权激励、风险补偿等多元政策组合，为企业成为创新主体提供便利。

3. 探索金融对科技创新的支撑作用

上海作为国际金融中心，探索金融对科技创新的支撑作用责无旁贷。一是上海需要积极加强多层次的资本市场建设，用好科创板、数据交易所、产权交易所等金融平台的作用，为企业的创新活动、成果转化和创新交易，提供多元化的金融服务。二是上海金融中心需要积极培育风险资本市场，打造中国版“纳斯达克”。要集聚和发展各类风

险资本，为企业创新、个人创业提供“种子资金”和“天使投资”，建立金融扶持实体经济的良好环境，全面激发各类市场主体的创新创业活力。三是金融资源应主动向创新领域倾斜。配合供给侧结构性改革、国家重点试验室、国家重大科技攻关项目等一系列重大工程、重点项目，政府应鼓励金融资源主动与之对接，并提供知识产权质押、股权融资服务、产权保护等系列配套服务，为企业专注研发提供良好保障。四是积极探索银团贷款、投资联动、保险增信等金融协同服务机制。推动直接融资与间接融资的协同，推动金融机构的主动合作，形成强大的金融支撑合力。

（五）上海科创中心建设要大力发展场景产业

1. 上海科创中心必须抓住人工智能的应用门户

人工智能目前仍处于发展的早期阶段，整体看来技术的发展将先于应用层面，但技术层面仍存在瓶颈需要突破；应用场景将不断丰富，它的扩充将会反过来驱动支撑技术的持续发展，人工智能的整体市场规模将继续扩大。在市场规模方面，综合考虑中国人工智能的爆发节点、技术成熟度以及全球人工智能市场规模等因素，我们保守估计2022年中国人工智能的市场规模应在700亿元左右。

在人工智能竞争的早期阶段，竞争主要围绕两个方面展开，一是争夺未来人工智能的门户接口，包括智能家居的平台接口、智能汽车的自动加强入口等，这些智能接口

担负着场景交互的重任，其功能类似于互联网时代的门户网站，也是各国和各大互联网企业集中争夺的领域。如谷歌的语音交互，百度的百度大脑。二是智能生态系统的竞争。智能生态系统一旦形成，平台用户就面临着极高的平台转换成本，因此各地区、各平台都在努力打造平台生态体系，以此形成稳固的用户群体。

上述两个竞争的重点，也为上海科创中心抓住新科技革命的机遇，提供了参考思路。尽管人工智能仍处于发展的早期，未来的产业形态、技术模式都还存在的较大的不确定性。因此，政府部门也就很难制定出清晰的技术路线图和产业发展规划，更多需要通过市场的竞争，通过大浪淘沙筛选出具有竞争优势的企业。与此同时，因为智能门户接口和智能生态两项，是人工智能革命在产业化落地过程中必然要抢占的高地，那么上海科创中心的建设就需要在突破原创性技术创新的同时，积极地去抢占这两个人工智能场景应用的关键环节。

2. 围绕场景产业打造科创中心的环境生态

相比应用场景层面有两项明确的关键点，上海科创中心建设在人工智能的基础层和技术层，所面临的挑战更为复杂。第一，人工智能仍处于早期，虽然语音识别、计算机视觉等感知层的通用技术目前已经取得了一定成就，主要的技术和成果已被国际大公司使用，但国内的北京、安

徽等省市已经取得技术领先，上海是否要跟进，需要认真考虑。虽然在通用技术层面留给后来者的技术空间和市场空间已经十分有限，上海想要介入上述领域，单纯从技术上看已经不具优势。但是如果从元宇宙的入口空间竞争角度看，这些通用技术仍极具战略价值，上海作为全球科创中心，仍有大力发展的必要。

第二，场景应用技术仍存在巨大的空间。随着人工智能支撑技术的不断发展，人工智能的场景应用也将持续拓展。多样化的应用场景，会产生众多具体的细分需求，需求的多样化会驱动支撑技术的完善和发展，从而带动整个人工智能行业的繁荣。但这种垂直领域的定制化应用场景，必然会被通用的智能平台所替代，这为上海科创中心研发通用型人工智能，并以此整合市场提供了机遇。

第三，上海具有人工智能场景落地的资本优势和市场优势。中国的人工智能应用有着鲜明的市场化特征，无法落地的技术很难得到资本青睐和市场认可。因此，上海科创中心的建设需要重点发挥好资本对技术研发和市场应用的支撑作用，利用好大资本的优势，推进人工智能技术与实际应用场景深度融合，以此形成开放技术体系下的不可替代性，进而产生难以被替代的商业价值。

第四，抓紧研发通用型人工智能。目前人工智能的应用和落地方式还极其有限，几乎所有人工智能都采用着输

入数据，然后快速生成简单回应的模式，这还只是人工智能的初级应用。上海科创中心需要紧紧抓住人工智能与新一代的5G、6G网络深度融合的机遇，把科技创新的重点放到提升整体生产、生产体系的应用效率上，着力解决中国制造转型升级的瓶颈问题，让人工智能成为产业升级的强大助力，改变社会生活的各个领域。

（六）大力推动长三角科创一体化发展

1. 加速打造区域协同创新共同体

科技协同创新是区域产业创新的根基和源泉。近年来，长三角区域的创新环境快速改善，创新资源、创新产出、产业潜力和环境支撑等一级指标显著提升，初具硅谷起步期的资源丰度和市场深度，长三角有望率先成为亚洲的科技创新中心。进一步完善长三角区域创新联盟，打造区域协同创新共同体，有利于激活区域创新活力，增强要素流通性，有利于上海发挥大都市的整体优势，从产业链的前后两端发力，整合区域研发设计和市场营销等创新资源，实现需求驱动与供给创新的双轮驱动，推动区域发展向创新驱动跨越。

2. 加快打造区域创新产业集群，构建区域现代产业体系

推动长三角科创一体化发展，做大做强创新型产业集群，能够放大市场规模，有利于上海构建现代化的产业体

系。多年来，上海一直面临着“创新成果落地难”“墙内开花、墙外结果”的困扰。上海的技术研发和创新成果，常常受制于高昂的商务成本和有限的土地空间，最终选择落户周边省份，导致前期高昂的研发投入难以回收成本，极大地抑制了上海的科技创新。

对此，上海在生物医药领域推出了“张江研发＋上海制造”行动，优化完善上海“1+5+X”生物医药产业基地布局，把握“研发＋临床＋制造＋应用”的全产业链，以加速创新成果在沪落地。生物医药产业的“张江模式”探索了全新的一体化产业组织模式，为上海科创中心探索了平台经济的新模式。未来上海可以以此为参考，进一步加强创新战略和创新政策的结合，联手长江流域各省协同推进体制机制创新，重点强化长三角区域的现代产业体系建设，以网络协同、利益共享等方式，推动区域科技创新与产业集群的一体化深度融合。

3. 发挥上海城市基础设施完善的优势，以产城融合吸引长三角创新资源

科技创新更加注重配套体系的完善，科创中心正从传统的“产业园区”模式，向具有城市综合功能、开放性和活力化的“城区”模式转变，这是上海集聚和整合长三角科创资源的优势和信心所在。

上海可以进一步推进创新功能空间与城市服务空间的

开放融合，在现有科技园区的基础上，将创新区域区分为“中心城区的创新区域、城市更新区域的创新区域、城市科技园（科学城）”三种类型。通过创新空间的细分和创新目标的细分，精准定位创新人才的客户群体，分类别、分特色构筑“科技创新区域”，并对口提供创新的服务支持，以满足各层次人才的创新创业需求，使上海科创中心真正成为优质人才、优质企业的创新之地、梦想之家。

附录

人工智能的场景应用

根据客户群体分类，人工智能的场景应用可以分为B2C应用和B2B应用两大类，目前正从分析型AI到生成式AI，不断逼近AGI。对标互联网的发展历程，B2C模式以及B2B模式在用户端的各向异性较弱，有利于更快积累数据，打造可持续竞争壁垒。因此，面向C端用户的场景应用，较面向企业B端的场景应用，在互联网的发展历程中得到了更多的重视，也发展得更早、更快。并且，先打通C端市场，有利于企业掌握终端消费用户的数据，实现对B端市场的场景包络，更有利于企业构建竞争优势。

于是，互联网企业普遍选择了率先完成B2C的场景应用，而B2B的场景应用则因为企业过于分散，数字技术和人工智能的应用相对缓慢。然而，在本轮新科技革命中，企业端的数字化、智能化因为以前积累有限，所以行业的竞争壁垒相对薄弱，反而有利于率先展开了人工智能的场景应用。只是因为B端距离公众较远，不为社会所熟知。

因此，我们场景应用案例也主要集中于企业。

根据场景应用功能分类，人工智能向各行各业的渗透与融合，基本是围绕着人工智能的五大基础能力展开，即图像生成、视频生成、语音生成、文本生成与代码生成。根据应用领域分类，五大基础能力孕育出丰富的应用场景，以 AI+ 的形式赋能各个行业。我们这里会简要介绍智能制造、智能教育、智能金融、智能农业、智能机器人、公共安全、智能医疗、智能传媒、智能商业等几个领域的典型场景应用，其中重点突出智能制造、智能教育的场景应用，以分别体现人工智能在制造业和服务业的典型应用。

一、人工智能技术应用的三个层面

人工智能技术与各行各业的加速融合，主要应用三个层面的核心能力，分别是计算智能、感知智能和认知智能。除了人工智能技术赋能各行各业的场景应用外，人工智能技术在这三个层面的核心能力，也得以直接应用于场景项目。

（一）计算智能的场景应用

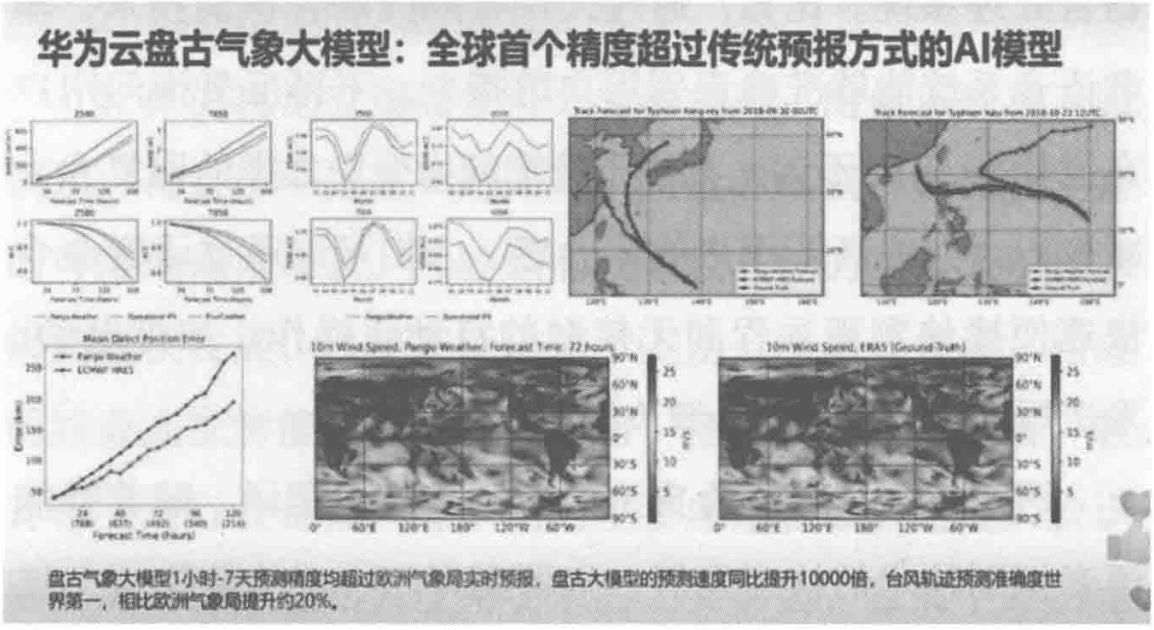
计算智能即机器具备超强的存储能力和超快的计算能力，可以基于海量数据进行深度学习，可利用历史经验指导当前的工作。计算智能的典型应用是阿尔法狗，2016 年 3

月，阿尔法狗与围棋世界冠军、韩国职业棋手李世石进行围棋人机大战，并以 4 比 1 的总比分获胜。阿尔法狗把计算智能的实力展现给世界，也开启了本轮人工智能的投资热潮。

计算智能的另一个典型应用是“AI+ 计算机”。计算智能可以自动生成程序代码，扮演一个虚拟程序员的角色。其典型的应用就是 OpenAI 公司的 ChatGPT。ChatGPT 可以根据精确或模糊的自然语言，了解用户的意图，并生成本文，写备忘录和诗歌，改写研究论文，解决数学问题，甚至编写软件。支撑 ChatGPT 强大功能的就是大语言模型的强大计算能力。其更为专业的场景应用则是借助 Debuild 自动编写程序，构建虚拟空间。用户只需说明自己想要的内容，人工智能即可以整合百万级程序员的智慧，根据注释和对方法的声明（参数、返回值），补充完成代码，完成程序的设计和 Web 应用部署。

计算智能的另一个典型应用是智能气候分析和气象预报。众所周知，气象分析是一个需要海量数据运算的庞大工程，各国无一例外地均使用超级计算机来完成此项任务。人工智能的出现，大大提升了人类气象分析和预测的水平，比如 2023 年 7 月 6 日《自然》（*Nature*）发表了华为公司的盘古大数据模型，其将人工智能应用在大型对撞机中以解释宇宙等议题，气象预报的速度比传统方法快 1 万倍，1.4 秒就可以完成全球 24 小时的气象预报，对台风、寒潮、热

浪等极端天气的预报准确性大为提高，表现出了极高的效率和准确度。



华为云盘古气象大模型

资料来源：华为公司官网

（二）感知智能的场景应用

感知智能是指使机器具备视觉、听觉、触觉等感知能力，将收集的非结构化数据进行结构化，实现与人类的沟通和互动。随着人工智能技术的飞速发展，更多非结构化数据被重视和挖掘，语音、图像、视频、触点等感知智能的关键技术被一一攻克。目前感知智能已经能够达到甚至超越人类的感知水平，也因此被广泛应用于各种场景，其中典型的有语音识别、无人驾驶、智能巡检等。

1. 语音识别和语音助手

广义自然语言处理技术在智能客服领域早有应用，但

多数都是有语料库支持的逻辑问答系统。最新的人工智能支持的语音识别系统是能够独立理解完整语句含义的自然语言处理系统。比如，通过人工智能的语音识别技术，车载语音系统能够准确识别用户的指令，不再反复询问用户在说什么，“对不起，无法听懂您的指令”之类的回复大为减少，大大提高了用户的体验感。人们可以通过语音命令快速便捷地实现远程和无接触的自动化操作，如打电话、发送短信、查询天气、调节家居设备等。

语音识别技术更为典型的应用是机器翻译。随着智能语音识别技术的快速发展，机器翻译正在成为跨语言交流中的重要工具，专业化的语音翻译软件已经开始替代高昂的同传翻译岗位，翻译的准确度已经能够匹敌同传翻译，而且不知疲倦。比如科大讯飞的“讯飞听见”软件，已经正式成为2023年布达佩斯世界田径锦标赛的赛事供应商，为赛事提供自动翻译软件和硬件产品及服务。其依托自然语言处理、声纹识别、语音识别、翻译等核心技术，形成了以转文字、云会议、拍字幕、找翻译、智能硬件、行业解决方案等为核心的智慧办公系列产品和服务平台，可满足多场景、多终端、多形式的办公需求。翻译软件能够快速准确地将一种语言翻译成另一种语言，不仅使得跨国交流更加便捷高效，而且可以应用于文献翻译、跨境旅游等多种场合，极大地降低了全球化的交流门槛，为企业和个

人创造了更为便捷、更为友好的商业交流环境和价值交换空间。

2. 机器视觉的人脸识别应用

计算机视觉应用场景可分为两大类：图像识别和人脸识别。得益于中国庞大的人口、大量的可获得数据及丰富的可落地商用场景，人脸识别技术较其他的图像识别技术得到更快发展，已经广泛应用于安防、金融等领域，相应的行业人工智能产品也较为丰富，商业价值十分显著。教育和交通行业也是人脸识别的主要应用领域，通过人脸识别人员身份，是相关应用场景的主要功能，目前中国的地铁、机场、小区门禁系统正在普及人脸识别。

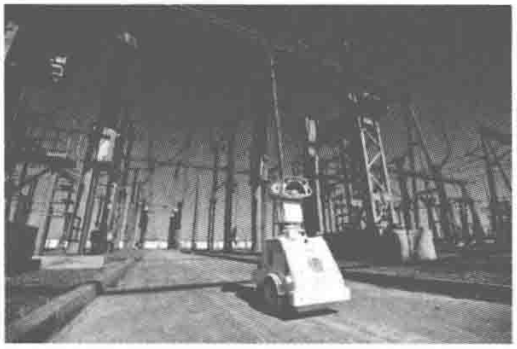
此外，人脸识别技术也正在向其他领域快速扩张，因其敏感性也引发巨大争议。比如娱乐营销场景是目前机器视觉发展最快的细分领域，各大视频网站纷纷采用人脸识别技术，以识别用户、分析喜好，以便布局视频广告以及潜在的内容审查应用。因此，用户很容易发现即使更换智能终端，更换上网地点，匿名登录某个视频 App，网站也会很快调整推送内容，给用户推送大致相同的视频，其主要机理就是 App 调用了智能摄像头的人脸识别功能。这种信息茧房现象被美其名曰用户体验的一致性，饱受各方诟病。

3. 机器视觉的图像识别应用

图像识别因为自然界的品种众多、数量巨大，更加依

赖人工智能技术和强大算力支撑，因此发展相对落后。目前为人熟知的应用场景是特斯拉汽车的自动驾驶项目所使用的图像识别系统，以及电力等系统的智能无人巡检系统等，我们重点介绍电力系统的巡检阅片。

电网巡检属于高风险、高强度、高技术的“三高”工作，始终面临着高压带电操作的危险，因此对机器替代人工的需求格外迫切。国家电网为此推出了“电力输电线路巡检图片智能识别系统”，巡检机器人代替人类进行定期的巡检工作。巡检机器人首先通过北斗系统进行高精度定位，并完成线路规划。其次，运用人工智能语音、图像识别、机器学习等技术，对变电站进行巡视、标记抄录和智能分析。再次，运用图像智能分析云服务，百张图像只需2分钟即可完成全部检测，实现海量数据的高效、精准分析。



智能电网的巡检机器人和智能无人机

资料来源：国家电网

以电力输电线路的绝缘子状态检测为例，主要任务有两项：一是目标检测；二是状态分类。目标检测的难度在于高压输电线路通常延伸数十甚至数百公里，检测需要花

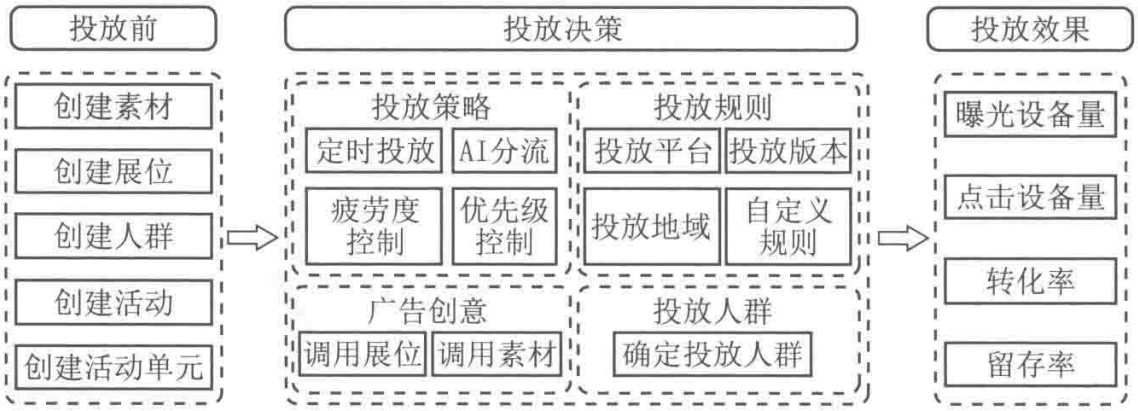
费较长时间，传统的人工检测需要翻山越岭，容易因疲劳等情况而产生漏检。而状态分类的难点在于，电线所处背景和气候非常复杂，野外摄像的图片分辨率不高，人工检测的误检率较高，查找故障点耗时较长。从 2010 年起，国家电网开始规模化应用直升机、无人机、可视化监拍装置和智能巡检机器人，实时传回输电线路状况，随之产生海量图片。以国网山东电力公司为例，2018 年该公司 600 架巡检无人机共传回了 120 万张图片。

对于这些状态图片，以往都是通过人工审图进行筛选和分析，手工对图片进行标注，每个电力班组一个小时可以审核 100 张左右的图片。以此为标准，完成山东分公司 120 万张的图片审核，需要 10 个班组连续工作 1000 多小时（约一年的工作时间），人员工作负荷过大，企业的成本开支也过大。2016 年国家电网公司组织开展了人工智能图像识别技术在输电线路巡检图像智能分析领域的技术攻关，用机器视觉来识别海量图片，具备了对导地线、绝缘子、异物、线夹及均压环、防振锤、细小金具、基础、附属设施、通道环境九大类缺陷识别的能力，可对 94% 的输电线路常见缺陷进行智能检测，直升机巡检作业销钉缺失检出率达到 89%，即便是对细小金属、线路异物等高难度缺陷，识别率也非常高。百张图像只需 2 分钟即可完成全部检测，效率是以前人工检测的数十倍。

（三）认知智能的场景应用

认知智能是指人工智能模拟人类的理解能力、归纳能力、推理能力，以及知识运用的能力。认知智能是人工智能技术的核心，目前还处于研究探索阶段。早在20世纪，公共安全领域就开始应用认知智能技术，对罪犯的行为进行特征提取和模式分析，开发出犯罪预测、资金穿透、犯罪模拟等人工智能模型和系统。在金融领域，人工智能技术既被用于宏观经济的预测分析，也被用于微观的可疑交易识别。但总体而言，由于认知智能的复杂性，目前距离真正的实用化尚有一定距离。

目前，认知智能商业化的典型案例之一是商品和服务的智能推送。比如，购物网站会根据用户的购买记录、浏览习惯，运用大数据模型推算出用户的兴趣爱好、可能购物需求等，并将相关商品放置到页面的显著位置，定向为用户推荐其可能购买的商品。而后台的供应链系统，则在



智能广告推送的决策流程

智能物流系统的算法调度下，一方面调度距离用户最近的库存，另一方面推算最优库存，以实现成本和效益的统一。

抖音是认知智能应用的成功案例。抖音风靡全球，主要得益于其强大的算法推荐机制。2020 年 8 月，抖音的日活跃用户已经超过 6 亿，每天都有几千万甚至上亿的视频被发布。面对如此巨大的数据量，平台首先需要审核是否为原创作品，是否存在抄袭、暴露等违法违规问题，这都需要通过人工智能进行鉴别。更为关键的是抖音如何通过算法，把这些视频精准地推荐给对此感兴趣的人。根据目前公开的信息，抖音应用了最新的认知智能技术，其中至少包括三个关键机制：一是智能推荐算法，包括推荐量的阶梯机制、推荐的叠加算法以及推荐的评估。二是智能推荐算法的五个核心权重指标，包括完播率、评论率、转发率、点赞率、复播率。三是抖音的系统审核机制，包括流程审核和双重审核等机制。通过这些复杂的评估模型和推荐算法，抖音借助大数据人工智能引擎和机器学习技术，能够快速地理解用户的喜好特征和兴趣点，精准描绘出用户的消费画像，并根据用户的行为习惯，推荐最符合其个性和兴趣的视频。

抖音这种基于推荐算法和认知智能的信息推送服务，实现了用户的个性化定制服务，可以满足不同类型用户的需求，大大提高用户黏性和平台忠诚度。而从用户角度来

看，抖音能够帮助自己快速找到需要的商品和服务，能够节省时间和精力，因此也并不会过度关注信息茧房的问题，反而深陷其中。

二、智能制造

近年来，随着以人工智能、互联网、大数据为代表的新一代信息技术加速与制造业深度融合，智能制造行业取得了长足发展。一方面，以“无人工厂”和“数字孪生工厂”为代表，“AI+ 制造”的应用场景日益丰富，人工智能已经能够贯穿于设计、生产、管理和服务等智能工厂的各个环节。另一方面，以工业机器人、高档数控机床为代表的关键技术装备产量快速增长，一场以工业添“脑”、产品换“芯”为特色的装备智能化，正在悄然推进。

当前，智能制造企业主要围绕智能语音、机器学习、人机交互、自然语言处理、生物特征识别、模式识别等多种人工智能技术，基本涵盖了计算智能、感知智能和认知智能的各个层面。主要的应用场景包括：智能分拣、设备管理、视觉检测、声纹检测、智能决策、需求预测、数字孪生、创成设计等。

（一）智能分拣

制造业可以分为流程型制造业和离散型制造业，离散

型制造业的产品由多个零部件组装而成，每个零部件在工厂运转的过程中是相对独立的，在众多的生产环节都存在一个分拣作业和质量检查的问题。比如，这些独立的零部件通常造型各异，并不会被整齐摆放，需要分拣作业将其正确归类和摆放。如果单纯依靠人工作业，速度过于缓慢且成本高昂。汽车有 1600 多个零部件，飞机更是有高达 20 万个零部件，单纯依靠人力分拣显然无法完成任务。现有的自动化生产流水线，工业机器人的摄像头即便识别了零件，也无法保证机器人每一次都按相同的位置成功抓取该零部件，或者机器人抓取该零部件后，还需要经过多次的动作调整，才能把不规则的零部件以相同的位置、角度、工序组装到产品上去。因此，面对生产制造过程中的这一现实需求，智能分拣技术应运而生。

亚马逊“全自动移动机器人”Proteus

以亚马逊的首款“全自动移动机器人”Proteus为例，该智能机器人于2022年6月推出，其应用场景为仓库物流系统，实质就是一个智能化的仓库搬运推车，能够在货物堆场更及时、准确地分拣货物，尽可能减少货物分拣过程中的搬运次数，保证货物的安全。

该移动机器人拥有“先进的安全、感知和导航技术”，人工智能的反复训练使机器人能够准确理解操作指令并对周围环境做出反应。智能分拣首先要解决机器人自身位置、货物位置和需要堆放位置等空间定位问题，借助SLAM（simultaneous localization and mapping，同步定位与地图构建）技术，自主移动机器人可以利用自身携带的传感器识别未知环境中的特征标志，然后根据机器人与特征标志之间的相对位置和以里程计的读数估计机器人和特征标志的全局坐标。这其中使用的主要是无人驾驶技术，需要应用多种人工智能技术与算法模型，解决空间定位、环境感知、路径规划、行为决策与行为控制等活动。

Proteus的应用场景还是生产线中相对较为简单的仓库物料供应环节。未来，随着人工智能技术的升级，智能分拣机器人必然更多地应用于操作更频繁、更复杂的生产流水线环节。

（二）设备健康管理

生产制造企业的设备众多，资产管理始终是难题。尤

其是流程型生产制造企业需要连续生产，生产流程相对固定、彼此关联，后续工序的开始以前道工序完成为前提，任何微小的设备问题都会直接影响到生产流程的运转，因此设备的健康管理对制造企业而言尤其重要。比如钢铁冶炼行业，高炉重启的费用惊人，因此通常保持连续运转，具有转速高、功率大、技术密集、价格昂贵、无备机、检修周期长等特点，必须保证设备及其备品备件的质量良好和数量充足。

设备的管理需要主要解决两个问题：一是尽可能提前发现设备的故障，减少非计划性停机；二是一旦发现故障，能够迅速进行故障诊断、定位故障原因并提供相应的解决方案。为此，需要人工智能系统介入对设备进行预防性检测，以提前预测和及时发现可能发生故障的设备，并提醒工厂及时更换。同时，需要人工智能系统实时监测设备的运行，实现跨模态智能自动调度及反馈。

以宝武集团的“黑灯工厂”为例，该工厂设有设备远程智能运维中心，被称为宝山基地的“大脑”。该中心基于工业互联网对设备进行在线智能管理，实现了传统制造企业由设备故障触发被动维护机制向基于设备远程运维平台开展主动预测性设备维护的模式创新和转变。整个无人工厂的冷轧区域是国内机器人配置数量最多的生产线，有无人天车、机器人自动喷涂、智能点检、自动捞渣、自动拆

捆等多个工种的智能化设备，约 14 个种类、479 台套。运维中心的远程智能运维，大幅改善了员工作业安全水平，提升了作业效率。

宝钢“黑灯车间”的智慧操作室

（三）智能决策

1. 生产管理决策

智能制造企业的决策包括产品质量、运营管理、能耗管理、设备管理、订单管理等多个方面，彼此的配合协作体现了企业的管理水平，也直接影响到企业的生产经营绩效。

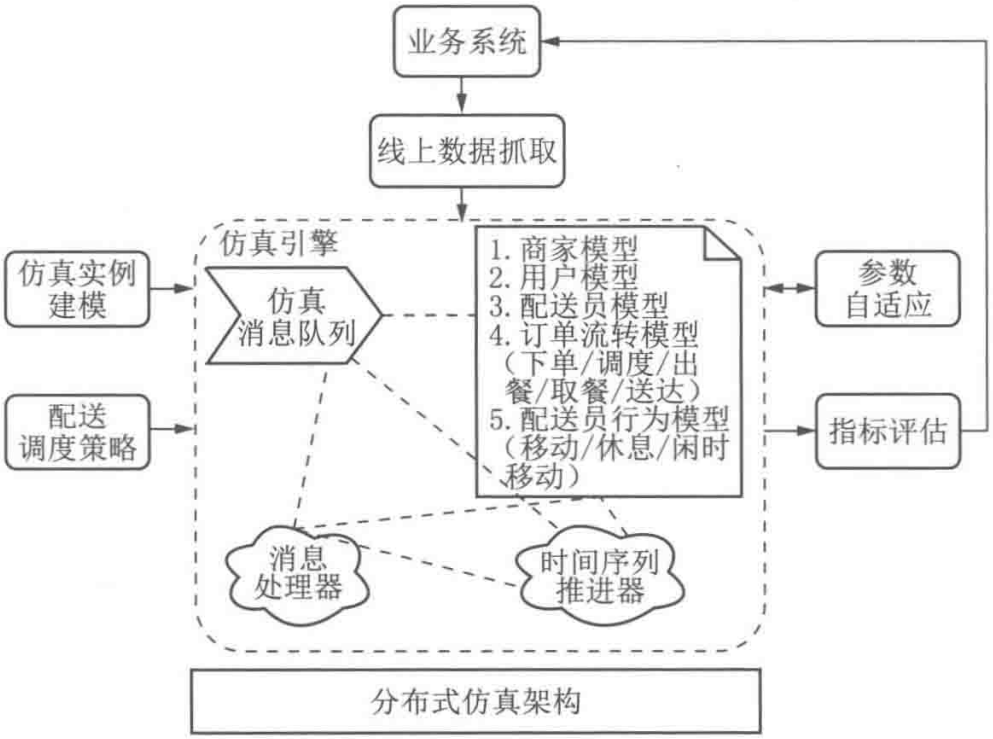
在生产管理决策方面，典型的场景应用案例是工信部推荐的“灯塔工厂”。以无锡柴油机厂为例，其是工信部首批智能制造试点示范项目单位，有 5 万平方米的生产区域，仅自动化总装线就有 179 个工位。其智能生产管理系统采

用了神经网络算法，通过将历史调度决策过程数据和调度执行后的实际生产性能指标作为训练数据集，建立大数据模型对调度决策评价算法的参数进行调优，完善了生产调度数据采集异常诊断、基于决策树的异常原因诊断、基于回归分析的设备停机时间预测、基于机器学习的调度决策优化等功能，对自主决策及调度响应速度进行秒级再优化，实现设备故障智能诊断、人员物料智能配置、刀具加工智能防错、能源消耗实时监控、动力设备自动启停、关键工序 100% 自动检测防错和数据分析应用，保证调度决策符合生产实际需求。

2. 订单分配决策

智能决策的另一个场景应用案例是订单分配，比如我们熟知的美团外卖，其给外卖骑手的订单分配就使用了智能决策技术。通过深度学习和数据分析等技术手段，智能订单分配可以快速地将订单分配给最适合的团队或人员，提高了订单的处理速度和准确性，进而提升了客户满意度，为商家和平台带来更多的商业价值。

智能订单系统的核心问题是如何实现即时配送。最直接的方法是增加人手，扩充运力规模，但配送成本会显著提高。因此，平台需要综合考虑配送体验和配送成本，在配送体验和配送成本之间取得最佳的平衡，这是整个智能订单系统的关键所在。



即时配送系统的分布式仿真架构

对此，美团推出了即时配送的“超级大脑”——O2O即时配送智能调度系统，其数据模型为带有若干复杂约束的DVRP（Dynamic Vehicle Routing Problem）问题，需要综合考虑骑手位置、在途订单情况、骑手能力、商家出餐、交付难度、天气状况、地理路况、未来单量等因素。通过对历史配送数据的分析和挖掘，会得到每个用户、楼宇、商家、骑手、地理区域的个性化信息，以及有关各地理区块骑行路径的有效数据，机器学习模块负责从数据中寻求规律和知识，为调度决策提供准确的基础信息。而运筹优化模块则在即时配送大数据平台以及机器学习的预测数据基础上，采用最优化理论、强化学习等优化策略进行计算，

做出全局最优的分配决策，并和骑手高效互动，处理执行过程中的问题，实现动态最优化。

三、智能教育

作为新一轮科技革命和产业革命的重要驱动力量，人工智能正在深刻改变人们的生活、工作和教育学习方式。尤其是2020年疫情以来，引发了一次史无前例的大规模在线教育，有效保障了近3亿师生的在线教学，也让人工智能、大数据、5G等新技术得以迅速走进教育体系。目前，智能教育产品已覆盖教、学、考、评、管、创新教育、五育并举等多个教育应用场景，推动整个教育行业率先开启了现代化的新征程。

人工智能在教育体系的场景应用可以粗略地概括为八个方面：智能辅导、微格教学、自适应学习、沉浸学习、自动测评、课堂评价、数据决策和教育治理。对学生而言，人工智能技术不仅为学生提供了智能化的教学产品和教学课程服务，使得知识获取的途径更加便捷，而且人工智能技术可以为学生量身定制学习计划、培养方案，使教学和培训更贴合学生的个性、潜力和兴趣，让学生的学习更加高效和快乐。对老师而言，智能教育系统和工具的引进，不仅能帮助老师自动整理和评审学生的作业、考

卷，让重复性劳动更加轻松，而且能够帮助老师自动整理学生的错题分类，形成汇总的分析报告和个性化的诊断报告，有助于老师因材施教、因人施训。因此，人工智能在教育领域的场景应用，在可预见的未来，将更加深入学生个性化学习的领域，有针对性地定制学习计划、提供教学方案。

（一）教学机器人

教育机器人本质是服务机器人，早在20世纪六七十年代，日本、美国、英国等西方发达国家已经相继在本国大学里开展了对机器人教育的研究，并推出了各自的教育机器人基础开发平台。中国在20世纪90年代开始初步发展教学机器人，直到疫情期间，在线教学的强劲需求终于拉动这个领域飞速发展。

教学机器人主要关注外观、听觉能力、视觉能力、辨识能力、口说能力、同理心与情绪侦测能力和长期互动能力七个领域。其中外观主要目的是识别学生，关注诸如脸型、体型、移动方式、性别、人格特质等。听觉能力主要关注语者辨识、语音识别、语义分析。视觉部分包括人脸侦测、人脸辨识、人脸追踪、姿势辨识、手势辨识、物体辨识、物体追踪。认人能力主要是解决多模辨识问题。同理心与情绪侦测能力主要包括情绪侦测、同理回应。长期互动能力包括长期记忆、持续行为、渐进行为。

十二类教育机器人产品应用场景

产品类型	说　明	产品案例
智能玩具	是一种可随身携带的电子零件且拥有智能的玩具。在满足儿童玩乐需求的基础上，加入教学设计，“寓教于乐”地引导儿童学习生活、语言、社交等知识。	Pleo、Dash-Dot
儿童娱乐教育同伴	针对 0 ～ 12 岁孩童而设计的同伴机器人。在家庭中，主要陪伴儿童学习，达到寓教于乐的效果。	爱乐优、Kibot-2
家庭智能助理	实体化为家庭智能助理的机器人可为个人解决家庭生活的问题，并提供相关服务。应用在教育上，可作为个性化学习服务的助理。	Jibo、Buddy、Pepper
远程控制机器人	使用者通过远程控制机器人，异地参与教或学的活动。	Engkey、Vgo
STEAM 教具	STEAM 是指科学、技术、工程、数学等多学科融合的综合教学方法，STEAM 教具则指根据 STEAM 教育理念设计的教学工具。	Mbot、Lego、Mindstorms、Ev3
特殊教育机器人	为特殊症状使用者设计的教学机器人。	Milo、Ask Nao
课堂机器人助教	协助教师完成课堂辅助性或重复性工作。	网龙华渔的未来教师
机器人教师	扮演教师角色，根据不同的教学情境，独自完成一门课程的教学，以达到教学效果。	日本东京理科大学 Saya 教师
工业制造培训	通过对企业内专业人员的培训，满足生产线的需求。	Baxter
手术医疗培训	培训医疗专业的工作人员，增加其对机器人手术操作的熟悉感。	达文西手术系统
复健照护	陪伴老年人专用的机器人，具备娱乐、脑力训练、复健教学等各方面复健照护的功能。	Zora Bot、Sil-bot
安全教育机器人	通过角色扮演，利用机器人传递安全教育的知识。	Robotronics

资料来源：网络

由上述内容可见，教学机器人基本涉及了人工智能的感知智能、认知智能和计算智能的各个领域，是服务机器人的集大成者。是人工智能、语音识别、人机交互、机器视觉、情境感知等技术应用在教育中应用的典型。从应用场景来看，教学机器人目前已经走出课堂，延伸到智能玩具、家庭智能助理等功能，以培养学生的分析能力、创造能力和实践能力为目标，未来将进一步向识别人才、创新人才的方向发展。

（二）课堂评价

课堂依旧是教育教学的主要时空场景，在线教育和网络课堂只不过是把现实的教学场景搬到了网上而已，并没有改变课堂的实质。

课堂评价的难点在于学生在课堂教学中始终是一个“黑箱”，老师无从判断学生接受知识的情况，只能通过课堂提问、课后作业和考试来判断大规模课堂教学的效果。人工智能的出现，使得可以通过数据采集、表情分析等技术，对课堂的情况进行分析判断。华东师范大学课程与教学研究所国际课堂分析实验室团队对多模态课堂分析进行了试验性探索，对课堂中的语音、图像、视频、空间、姿态以及多模态数据进行采集与分析，全方位追踪课堂轨迹，采用人工智能技术，提取与解析课堂多维度数据，进而实现智能诊断与智能反馈的一体化课堂教学与研究。采集的

信息包括学生举手次数、教师提问次数、学生回答次数、教师表情、学生表情等，目前已经形成了众多的研究成果和场景应用。

（三）VR 课堂

在技术层面，新兴技术也正在与在线教育领域积极结合，并催生出一些有趣的新教育模式。比如美国教育科技公司 Nearpod VR 已经向美国 1 万多所学校提供 VR 课程，展开了基于 VR 的线上课程实验。与 VR 课堂相适应，人工智能技术也被大量应用于在线教育，比如美国数学辅导软件公司 Carnegie Learning，2022 年推出了 Livelab 项目，用人工智能部分替代老师功能，为学生提供个性化的指导和评估。

VR 课堂只是人工智能技术应用于在线教育的一个典型代表，事实上美国已经在线上教育培训领域领跑许久，目前已经全面进入了细分教育的时代。机器人编码（Wonder Workshop）、创造力开发（Curious）、课堂学习过程追踪反馈（Skillshare）、在线摄影学习（Digital-photography-school）、在线音乐学习（ClassDojo）、科技产品 DIY（Makezine）、职业内容学习平台（Digital Tutors）等细分企业层出不穷，给在线教育注入了全新血液。根据市场研究机构 docebo 的研究显示，截至 2020 年 2 月，整个美国共有 27939 个关键词为“e-learning”的公司。

世纪出版

本书作者以开阔的视野，严密的逻辑和细致的文笔，十分可信地论述了中国现代化建设一定离不开新一轮科学技术革命（类脑技术革命）的加持。作者独具匠心地运用金字塔三层结构模型，阐述了人工智能技术在相关语音识别、语义识别、图像识别、神经网络、深度学习等一系列技术的融合创新及新基础设施之上呈现出的广泛应用前景，其必将成为未来人类幸福生活、新型生产、科技创新、社会运行与管理等不可或缺的核心。可以说这是一本给人以震撼与启迪的书，特此推荐。

复旦大学管理学院产业经济学系主任，
中国工业经济学会副理事长，一级教授　**芮明杰**

在我们受托编制和发布的上海2万多家高新技术企业中的2023年创新百强榜上，人工智能开发企业达到36家。这是人工智能正在迅猛发展的一个观照。本书给予我们人工智能引领新一轮科技革命的清晰图景，有思想、有观点、有前瞻、有对策。处在人工智能的时代性浪潮中的我们，认真读一读这本书很有必要、大有裨益。

上海社会科学院创新创业经济研究中心主任，
上海社会科学院应用经济研究所原副所长，教授　**李　湛**

上海市社会科学界联合会
Shanghai Federation of Social Science Associations

上架建议：人工智能

ISBN 978-7-208-1851

9 787208 185197 >

定价：72.00 元

易文网：www.ewen.co